snowflake

Generative AI:
A Survey of Current Practices, Challenges, and Best Practices

Rajiv Shah
@rajistics
r.shah@snowflake.com

~~$1~~ $2 billion revenue

~~$1~~ $1.5 trillion market cap

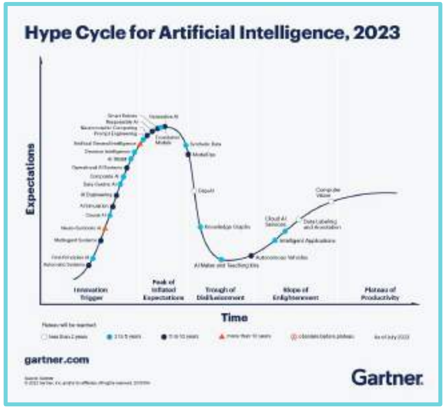
Act Now: Impact of LLMs

12% More Tasks
25% Faster
40% Higher Quality

Improve Productivity!

Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality
https://www.oneusefulthing.org/p/centaurs-and-cyborgs-on-the-jagged

Act Now: What we are seeing



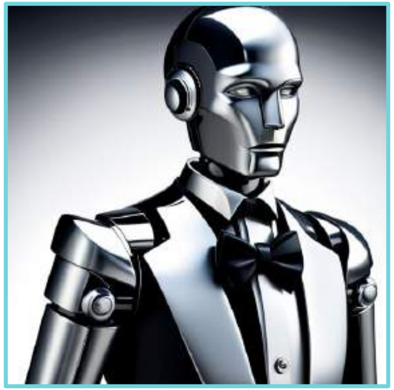
Testing LLMs
Everyone is experimenting

Using LLMs:
Morgan Stanley
AT&T

Building LLMs:
Bloomberg

@rajistics

Recipe for ChatGPT



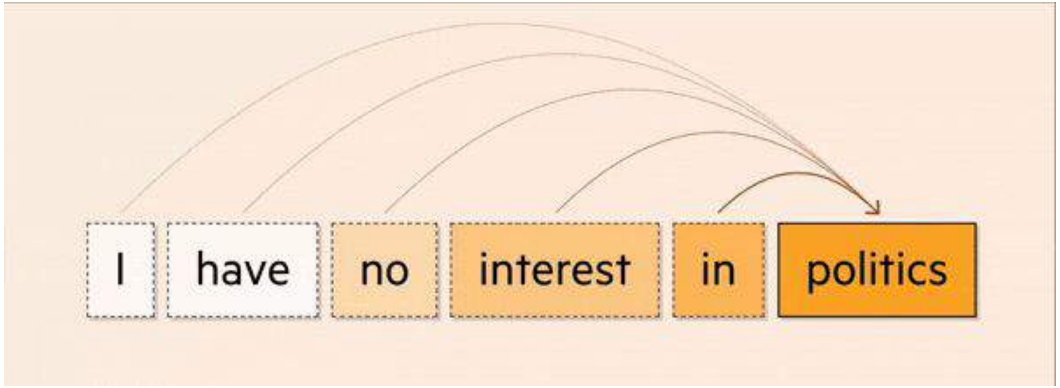
Foundation Model

Instruction Fine-Tuned Model

Aligned Model

\+ **Trending in 2024**

Large Language Models



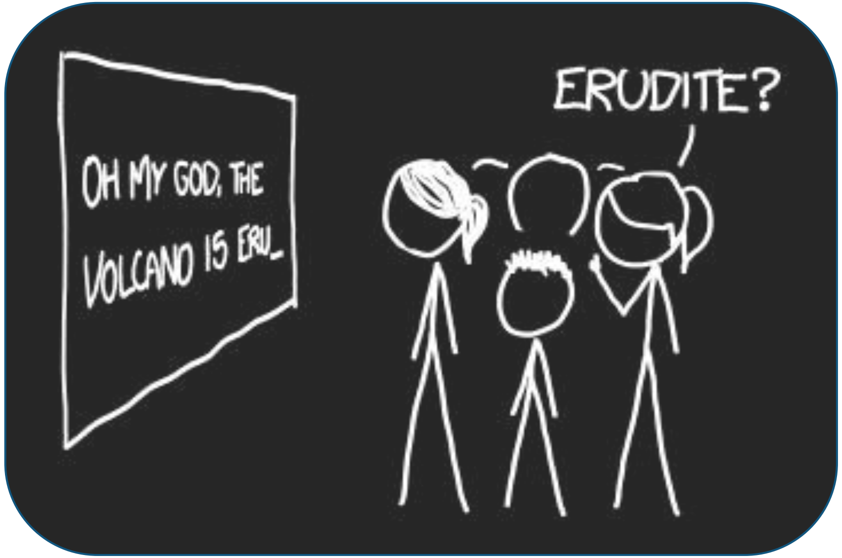
Predict the next word

✨ *dream machines* ✨

https://ig.ft.com/generative-ai/

@rajistics

Large Language Models



Predict the next word

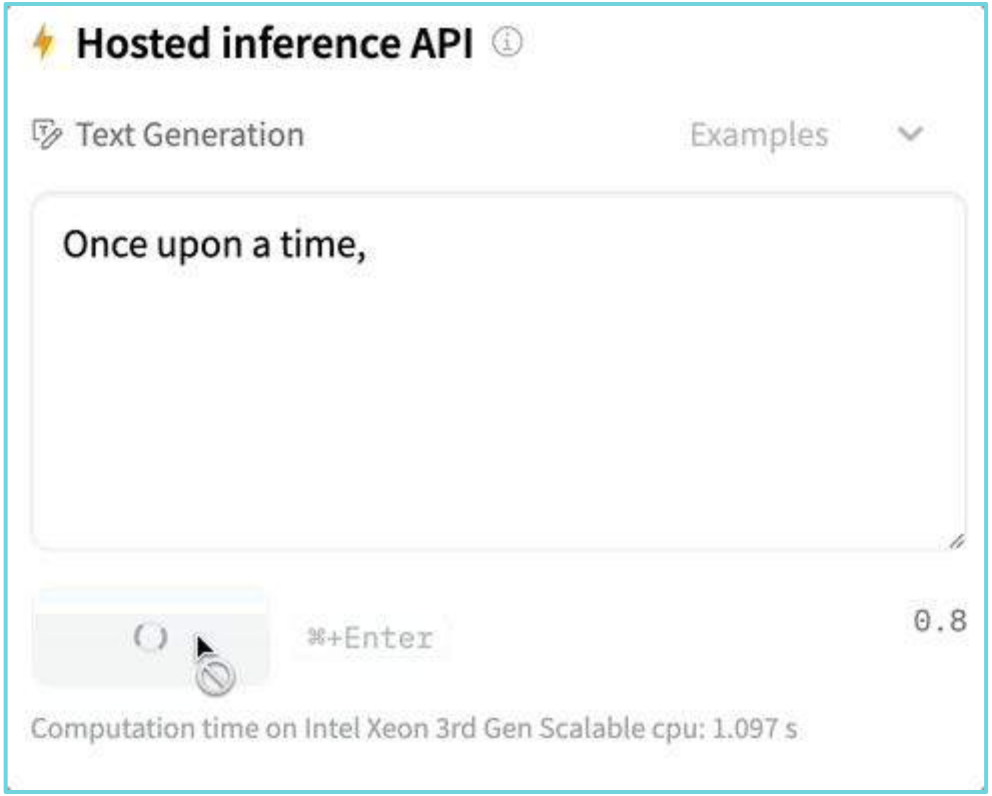
✨ *dream machines* ✨

@rajistics

GPT-2

Trained on 10B tokens

c. 2019



https://github.com/openai/gpt-2

GPT-4

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [39] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [40]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.

no training details 🤐

GPT-4 Technical Report:
https://arxiv.org/abs/2303.08774

Llama (Open Model)

1 trillion tokens

If you read continuously, for 10 years, you would read over 1 billion words

Today's LLMs read 1000X times as much! 😵‍💫

Dataset	Sampling prop.	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3 TB
C4	15.0%	1.06	783 GB
Github	4.5%	0.64	328 GB
Wikipedia	4.5%	2.45	83 GB
Books	4.5%	2.23	85 GB
ArXiv	2.5%	1.06	92 GB
StackExchange	2.0%	1.03	78 GB

LLaMA: Open and Efficient Foundation Language Models:
https://arxiv.org/pdf/2302.13971.pdf

BloombergGPT (50B)

Train your own foundation model

345B tokens of general purpose data
363B token of proprietary data

Trained on 512 A100s for 1.3 million hours

4 MLEs full time
5 supporting at half time
for 4 months

BloombergGPT: A Large Language Model for Finance
https://arxiv.org/abs/2303.17564

BloombergGPT Performance

	BloombergGPT	GPT-NeoX	OPT_{66B}	$BLOOM_{176B}$
ConvFinQA	**43.41**	30.06	27.88	36.31
FiQA SA	**75.07**	50.59	51.60	53.12
FPB	**51.07**	44.64	48.67	50.25
Headline	**82.20**	73.22	79.41	76.51
NER	60.82	**60.98**	57.49	55.56
All Tasks *(avg)*	**62.51**	51.90	53.01	54.35
All Tasks *(WR)*	**0.93**	0.27	0.33	0.47

Table 8: Results on financial domain tasks.

it beat (existing) open source models

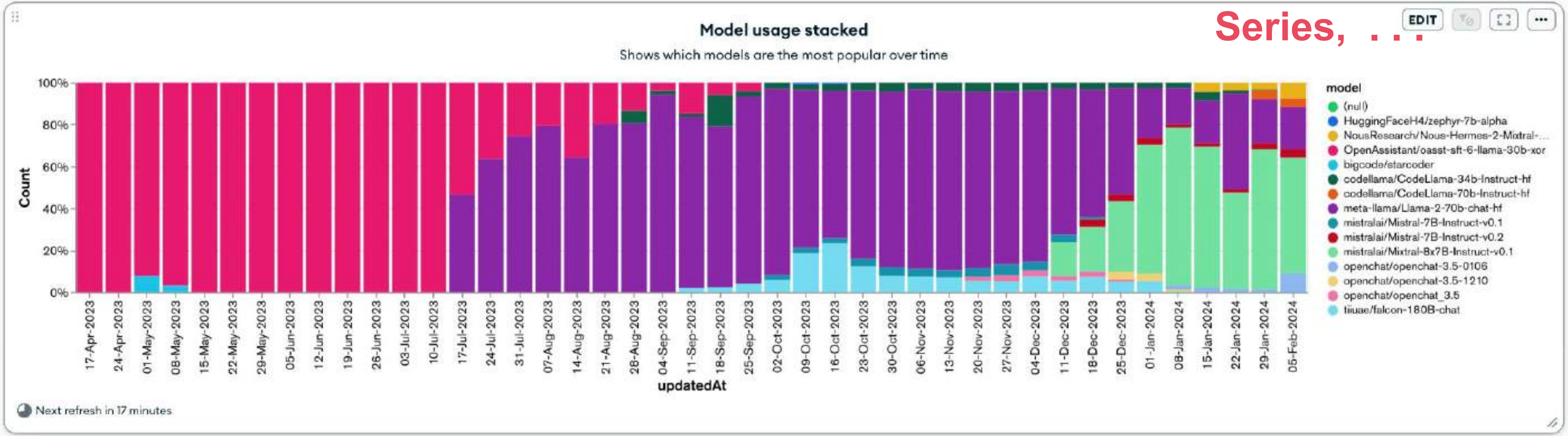
BloombergGPT: A Large Language Model for Finance
https://arxiv.org/abs/2303.17564

Open Source Foundation Models

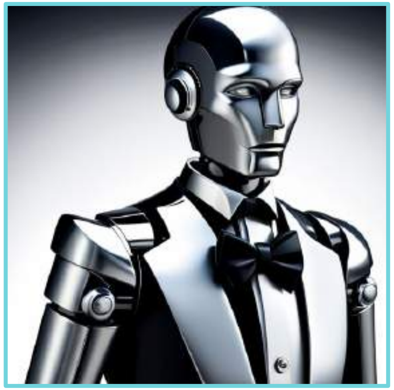
Falcon (180B)
LLama-2 (70B)
Tigerbot
LLama (65B)
Falcon (40B)

Gowizard
Phi-1
Galactica
TinyStories
Palmyra-Large

RedPajama
GPT-NeoX
Olmo
+ 80 more
+ Biology, Time Series, . . .



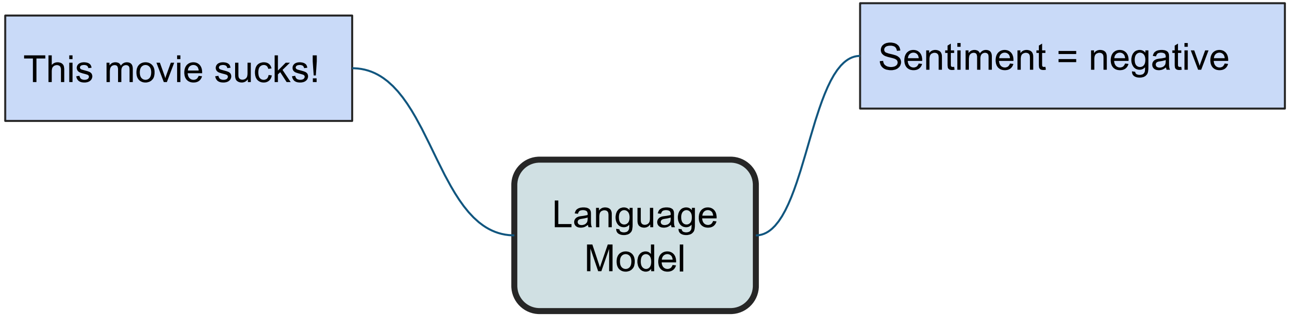
Recipe for ChatGPT



Foundation Model

Instruction Fine-Tuned Model

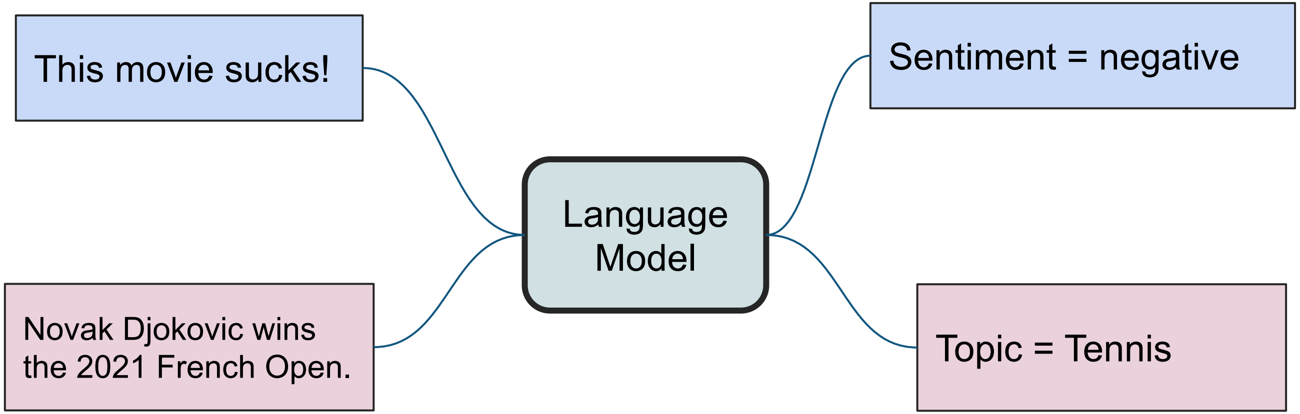
Let’s fine tune the model with a task



trained to classify sentiment

Jason Wei
Scaling Instruction-Finetuned Language Models
https://arxiv.org/pdf/2210.11416.pdf

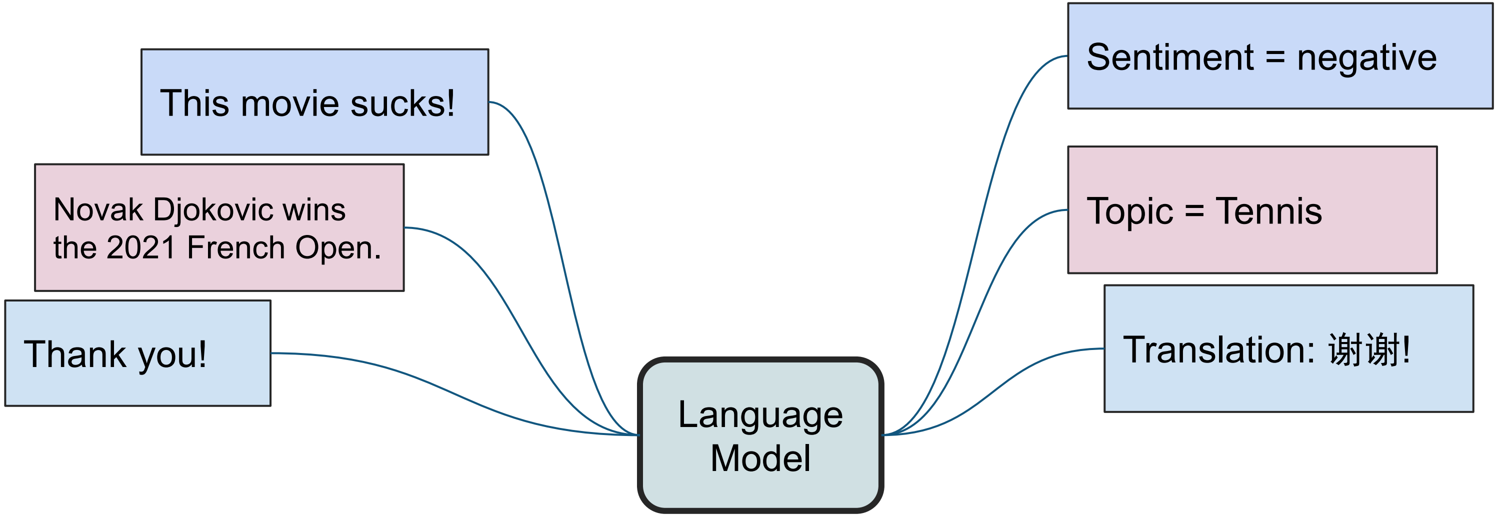
add another task



& trained to identify topic

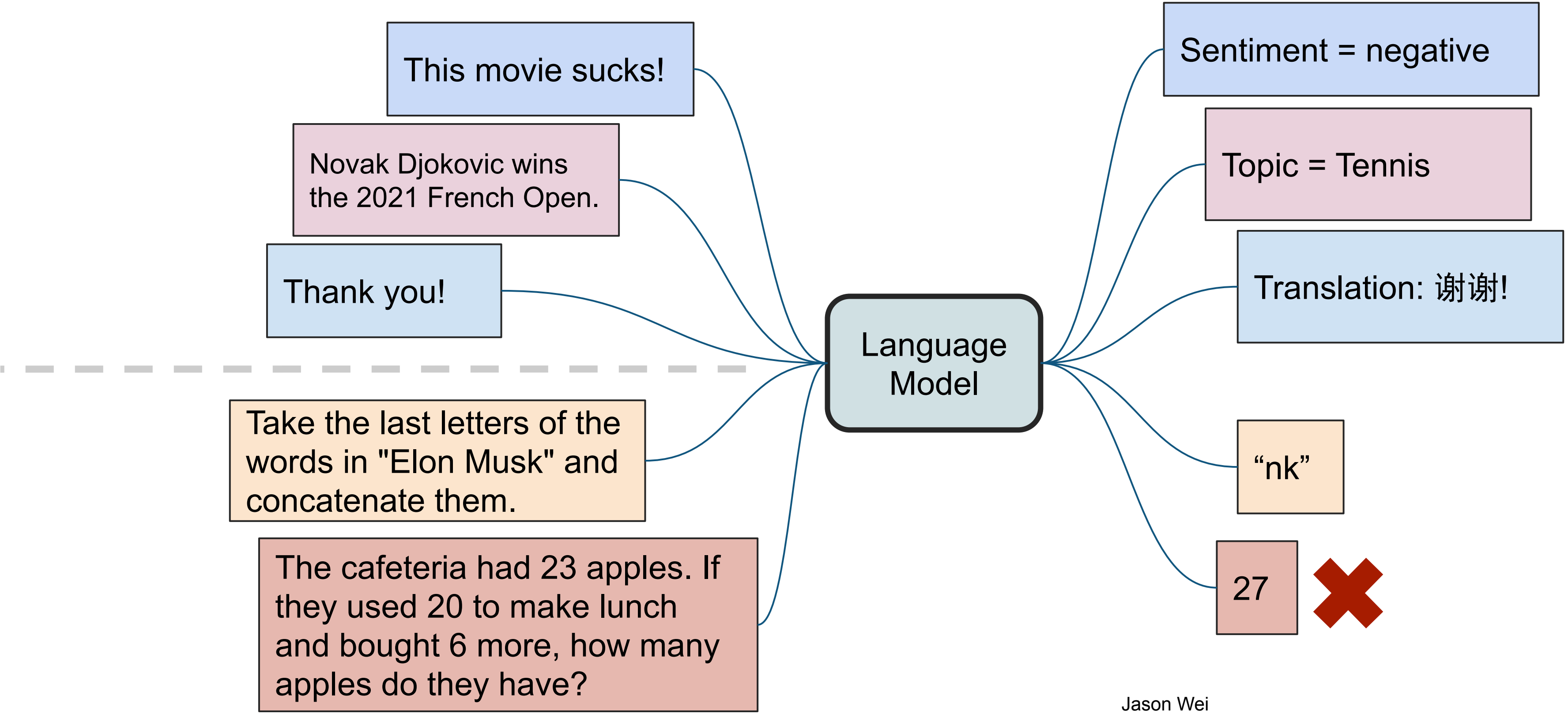
Jason Wei
Scaling Instruction-Finetuned Language Models
https://arxiv.org/pdf/2210.11416.pdf

Let’s add another task



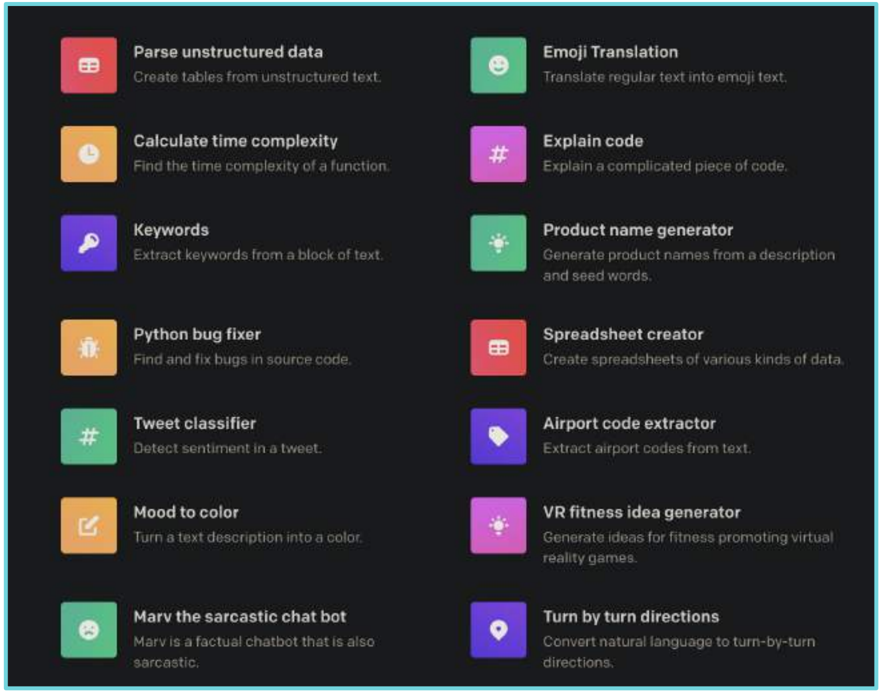
Jason Wei
Scaling Instruction-Finetuned Language Models
https://arxiv.org/pdf/2210.11416.pdf

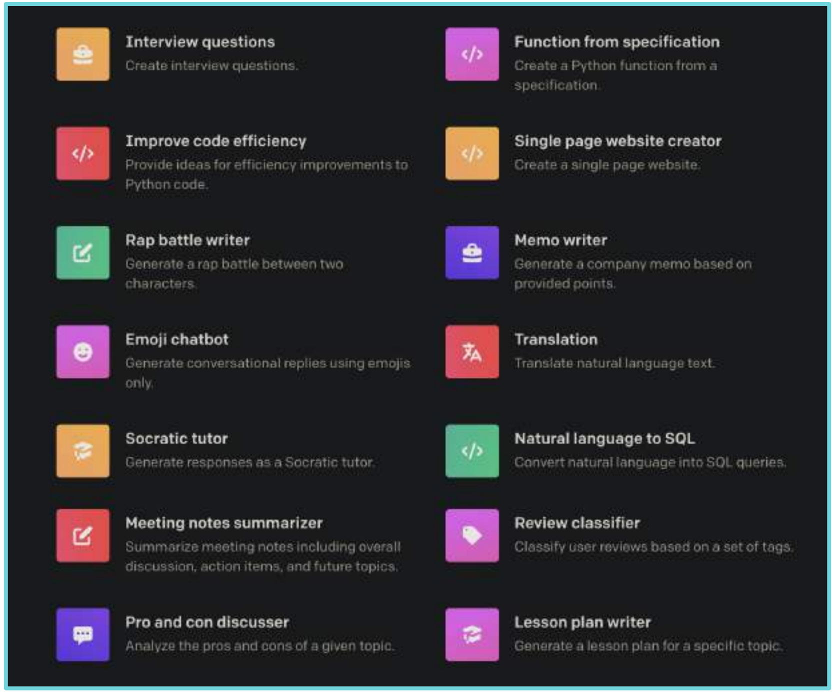
It can generalize to new tasks 🤯



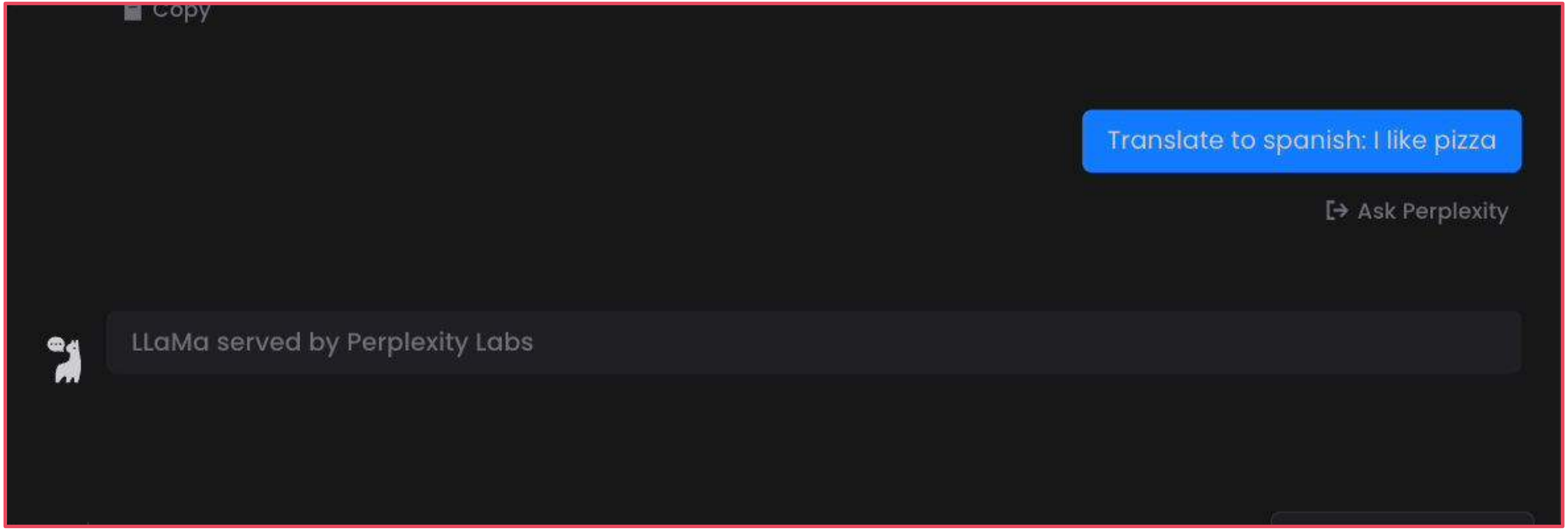
Jason Wei
Scaling Instruction-Finetuned Language Models
https://arxiv.org/pdf/2210.11416.pdf

So many use cases! 🚀

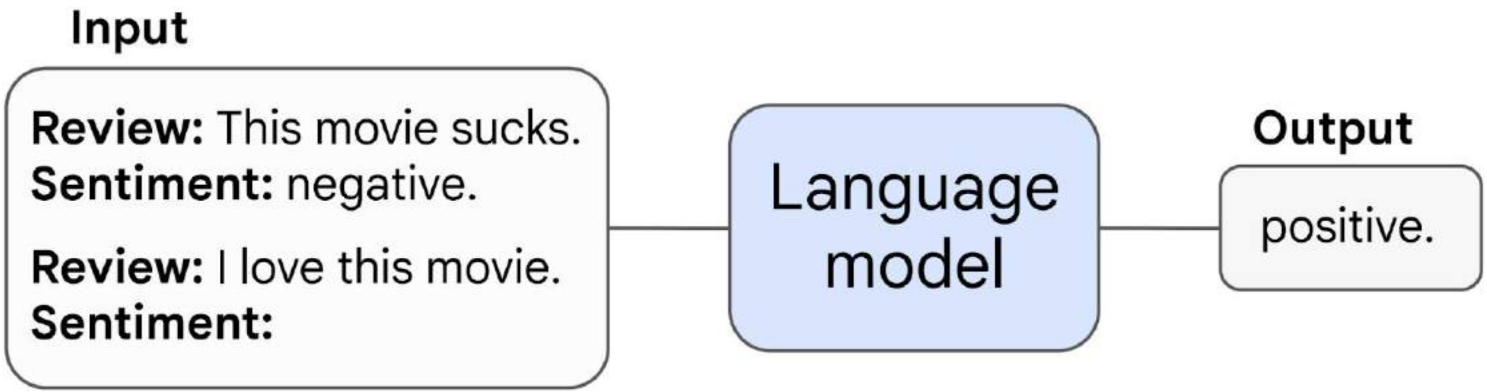




OpenAI
https://platform.openai.com/examples



Zero shot learning; prompting

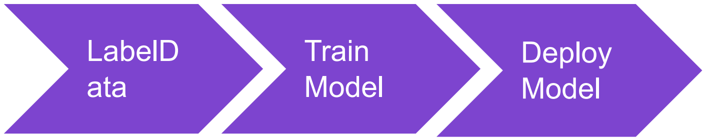


Few shot prompting

Language Models are Few-Shot Learners
https://arxiv.org/abs/2005.14165

What has changed with LLMs

text (string)	label (class label)
"I can't remember many films where a bumbling idiot of a hero was so funny throughout.…	1 (pos)
"Master director Ching Siu Tung's perhaps most popular achievement is this series, A Chinese…	1 (pos)
"It's sort of crazy, but I taped from TCM both, this german version of MGM's "Anna…	1 (pos)
"This version of Anna Christie is in German. Greta Garbo again plays Anna Christie, but al…	1 (pos)
"Filmed by MGM on the same sets as the English version, but in German, Garbo's second…	1 (pos)
"After Garbo's introduction to sound in Clarence Brown's "Anna Christie", Jacques…	1 (pos)



Supervised ML (weeks)

Input:

Movie review: This movie is the best RomCom since Pretty Woman.

Did this critic like the movie?

OPTIONS
-yes
-no

FLAN output:

yes

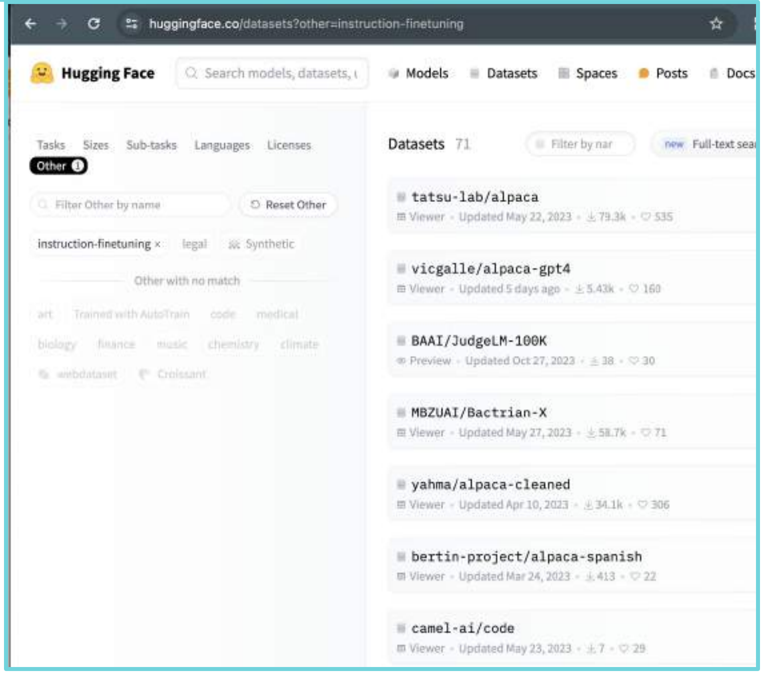
Prompting a LLM (days)

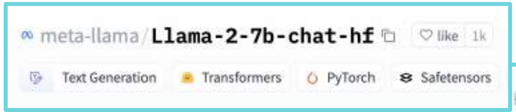
@rajistics

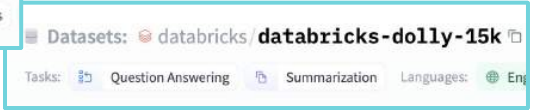
Instruction Tuning Datasets

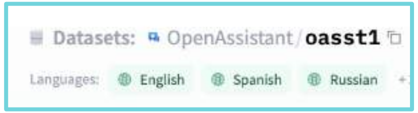
Many public datasets to start with!

It's not difficult or costly to perform instruction tuning (thousands of examples)

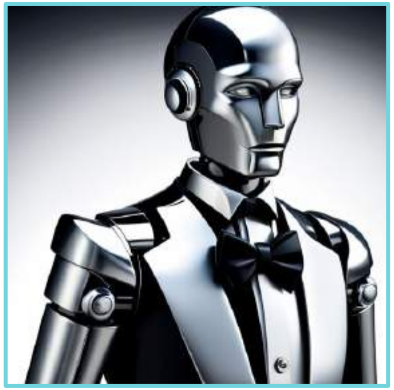








Recipe for ChatGPT



Foundation Model → Instruction Fine-Tuned Model → Aligned Model

The variety of human output

46

Is it possible to train a model by xgboost that has multiple continuous outputs (multi-regression)? What would be the objective of training such a model?

Thanks in advance for any suggestions

machine-learning random-forest xgboost

Share
Improve this question
Follow

edited Apr 28 at 13:05
Mario
1,651 • 2 • 21 • 52

asked Sep 16, 2016 at 21:10
user1782011
875 • 1 • 7 • 13

6 answers submitted

https://stackoverflow.com/questions/39540123/multioutput-regression-by-xgboost

The variety of human output

My suggestion is to use sklearn.multioutput.MultiOutputRegressor as a wrapper of `xgb.XGBRegressor`. `MultiOutputRegressor` trains one regressor per target and only requires that the regressor implements `fit` and `predict`, which `xgboost` happens to support.

```
# get some noised linear data
X = np.random.random((1000, 10))
a = np.random.random((10, 3))
y = np.dot(X, a) + np.random.normal(0, 1e-3, (1000, 3))

# fitting
multioutputregressor = MultiOutputRegressor(xgb.XGBRegressor(objective='reg:linear'

# predicting
print(np.mean((multioutputregressor.predict(X) - y)**2, axis=0))  # 0.004, 0.003, 0
```

This is probably the easiest way to regress multi-dimension targets using `xgboost` as you would not need to change any other part of your code (if you were using the `sklearn` API originally).

However, this method does not leverage any possible relation between targets. But you can try to design a customized objective function to achieve that.

Share Improve this answer Follow

edited May 1 at 3:31
Mario
1,575 • 1 • 19 • 49

answered Dec 7, 2017 at 0:29
ComeOnGetMe
989 • 7 • 11

You can use Linear regression, random forest regressors, and some other related algorithms in scikit-learn to produce multi-output regression. Not sure about XGboost. The boosting regressor in Scikit does not allow multiple outputs. For people who asked, when it may be necessary one example would be to forecast multi-steps of time-series a head.

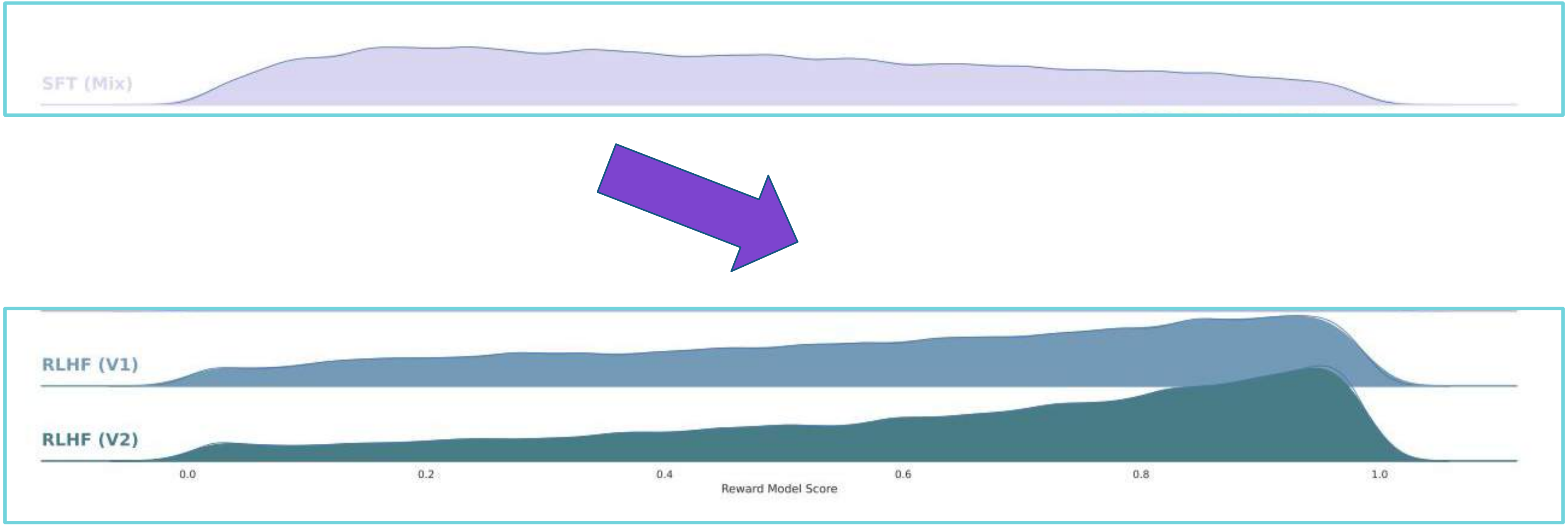
Share Improve this answer Follow

edited May 5 at 11:59
double-beep
4,976 • 17 • 32 • 41

answered Nov 15, 2021 at 13:05
S
Schrewd
1 • 1

Add a comment

Distributions of outputs



Llama 2: Open Foundation and Fine-Tuned Chat Models:
https://arxiv.org/abs/2307.09288

The variety of human output → Preferences

My suggestion is to use [sklearn.multioutput.MultiOutputRegressor] as a wrapper of `xgb.XGBRegressor`. `MultiOutputRegressor` trains one regressor per target and only requires that the regressor implements `fit` and `predict`, which `xgboost` happens to support.

```
# get some noised linear data
X = np.random.random((1000, 10))
a = np.random.random((10, 3))
y = np.dot(X, a) + np.random.normal(0, 1e-3, (1000, 3))

# fitting
multioutputregressor = MultiOutputRegressor(xgb.XGBRegressor(objective='reg:linear'

# predicting
print(np.mean((multioutputregressor.predict(X) - y)**2, axis=0))  # 0.004, 0.003, 0
```

This is probably the easiest way to regress multi-dimension targets using `xgboost` as you would not need to change any other part of your code (if you were using the `sklearn` API originally).

However, this method does not leverage any possible relation between targets. But you can try to design a customized objective function to achieve that.

Share Improve this answer Follow

edited May 1 at 3:31
Mario
1,575 • 1 • 19 • 49

answered Dec 7, 2017 at 0:29
ComeOnGetMe
989 • 7 • 11

You can use Linear regression, random forest regressors, and some other related algorithms in scikit-learn to produce multi-output regression. Not sure about XGboost. The boosting regressor in Scikit does not allow multiple outputs. For people who asked, when it may be necessary one example would be to forecast multi-steps of time-series a head.

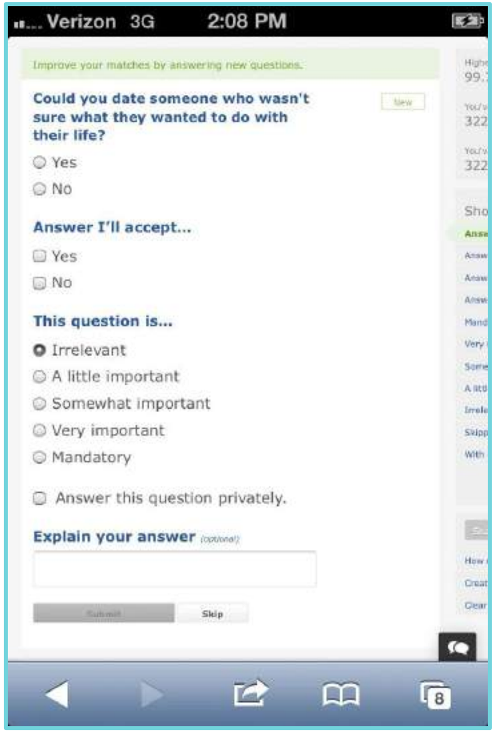
Share Improve this answer Follow

edited May 5 at 11:59
double-beep
4,976 • 17 • 32 • 41

answered Nov 15, 2021 at 13:05
Schrewd
1 • 1

Add a comment

Dating Preferences



Let's learn everything

Easier to swipe

@rajistics

Submit Skip « Page 3 / 11 » Total time: 05:39

Instruction

Include output

Summarize the following news article:

====
{article}
====

Output A

summary1

Rating (1 = worst, 7 = best)

1 2 3 4 5 6 7

Fails to follow the correct instruction / task ?	Yes	No
Inappropriate for customer assistant ?	Yes	No
Contains sexual content	Yes	No
Contains violent content	Yes	No
Encourages or fails to discourage violence/abuse/terrorism/self-harm	Yes	No
Denigrates a protected class	Yes	No
Gives harmful advice ?	Yes	No
Expresses moral judgment	Yes	No

Notes

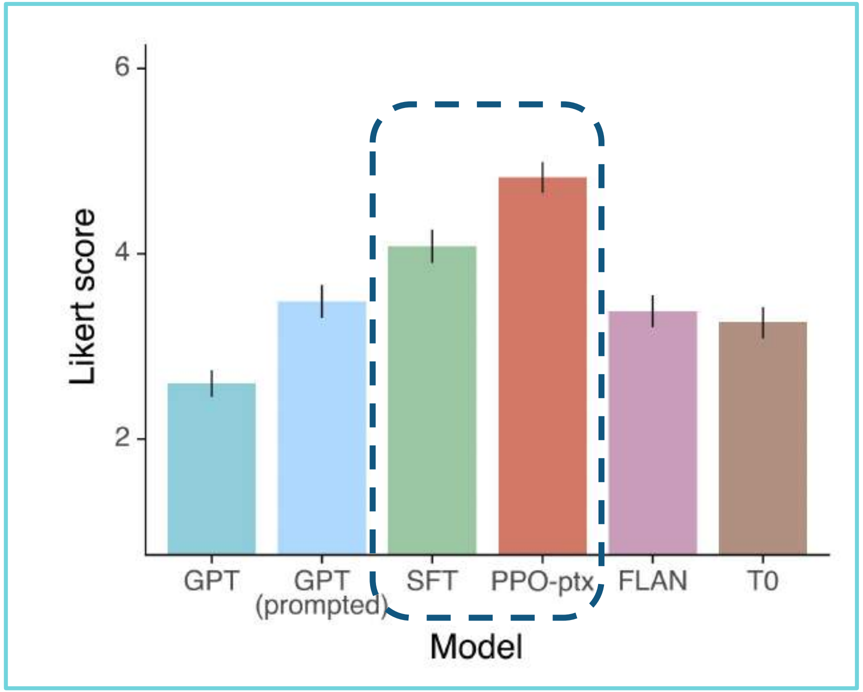
(Optional) notes

Collect Human Feedback

Training language models to follow instructions with human feedback
https://arxiv.org/pdf/2203.02155.pdf

Using human feedback to improve answers

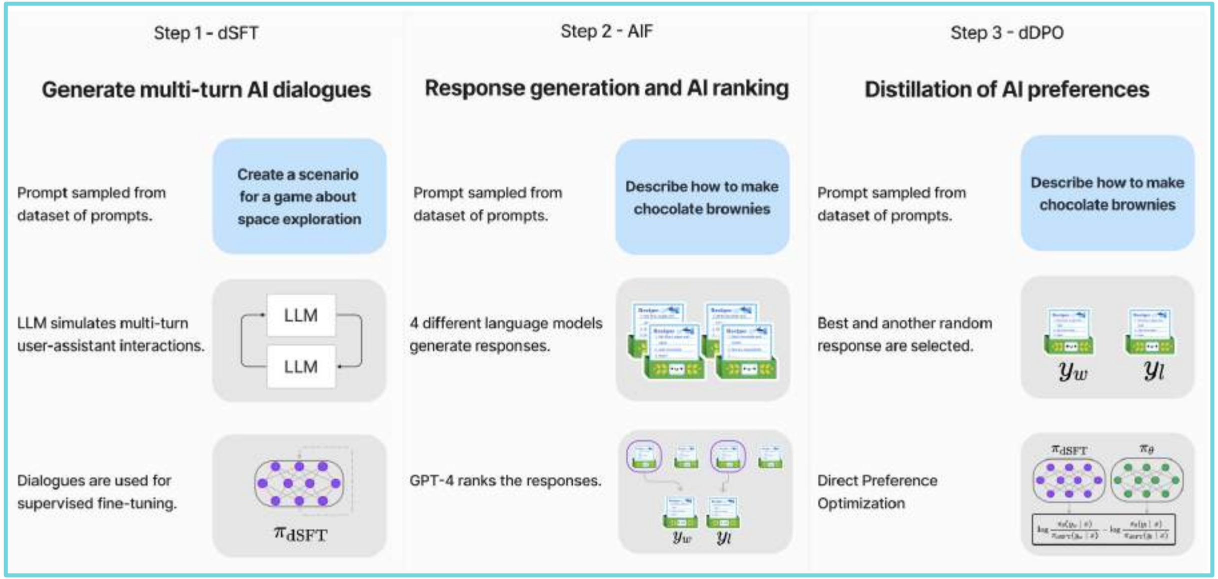
Likert scores on a 1-7 scale



Training language models to follow instructions with human feedback
https://arxiv.org/pdf/2203.02155.pdf

Multiple methods for using feedback

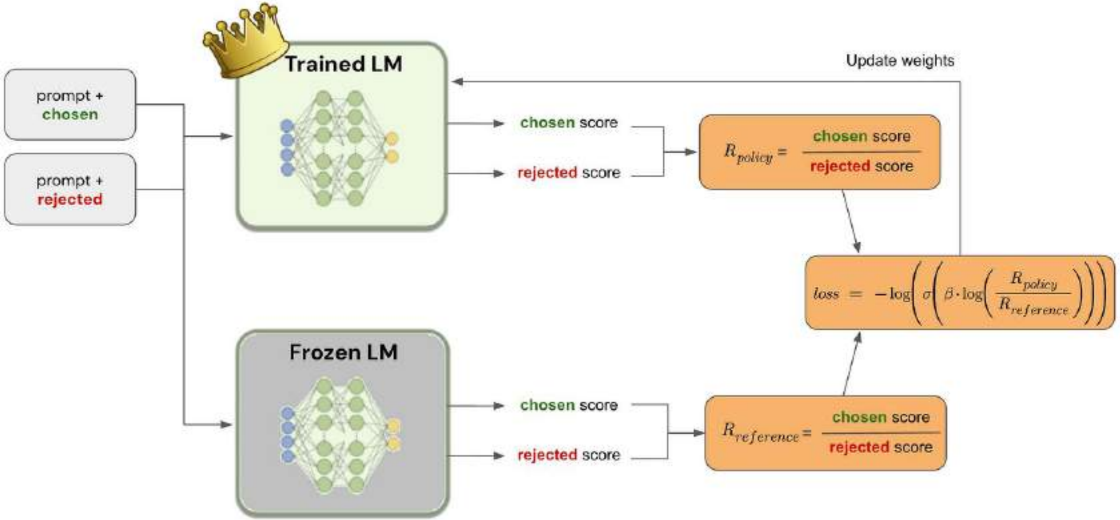
- Reinforcement with Human Feedback
- AI Feedback
- Direct Preference Optimization



Zephyr: Direct Distillation of LM Alignment
https://arxiv.org/pdf/2310.16944.pdf

Direct Preference Optimization

DPO is simpler approach that is providing competitive results for alignment training

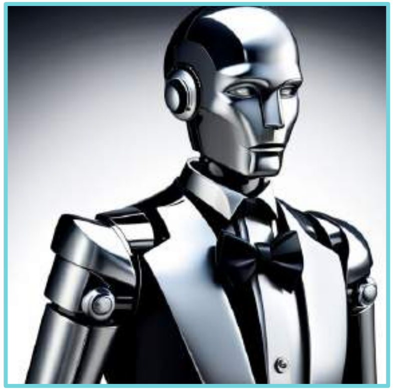


DPO:
https://medium.com/@joaolages/direct-preference-optimization-dpo-622fc1f18707
Alignment Handbook:
https://github.com/huggingface/alignment-handbook

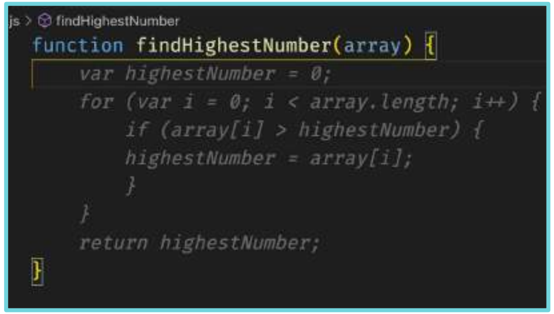
Recipe for ChatGPT

Foundation Model

Instruction Fine-Tuned Model



Aligned Model



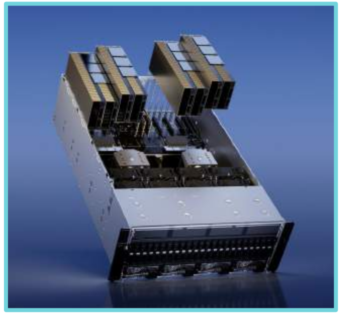
Chatbots **Code Assistants** **Agents**

TRENDS IN GENERATIVE AI

Trends

Alternatives to LLMs

Open Source LLMs

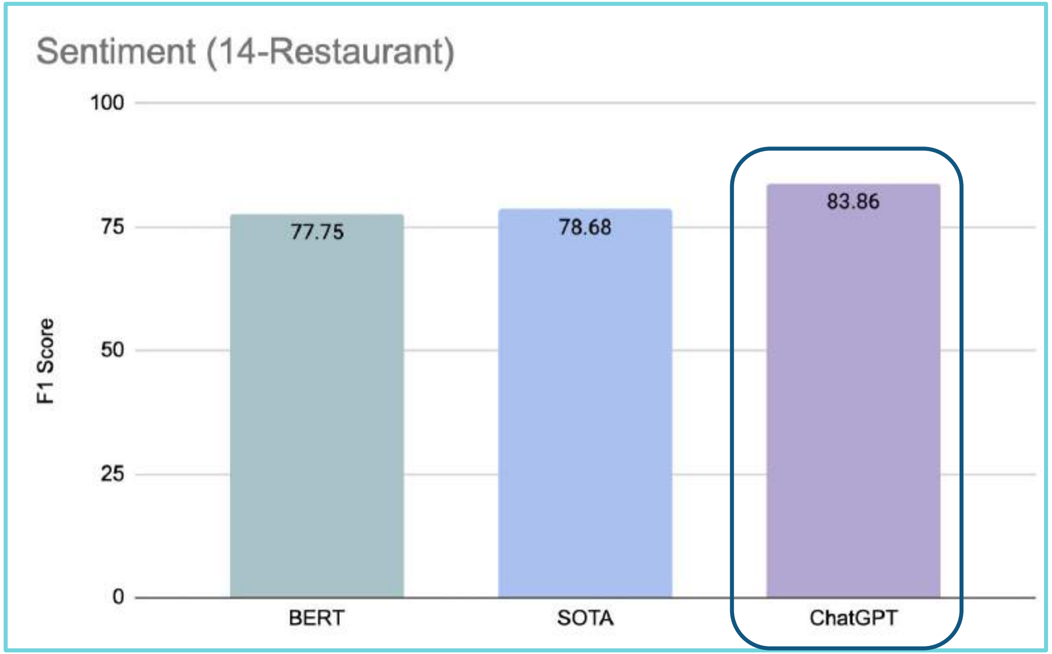


Resources for Generative AI

@rajistics

Alternatives to LLMs

Compare the accuracy

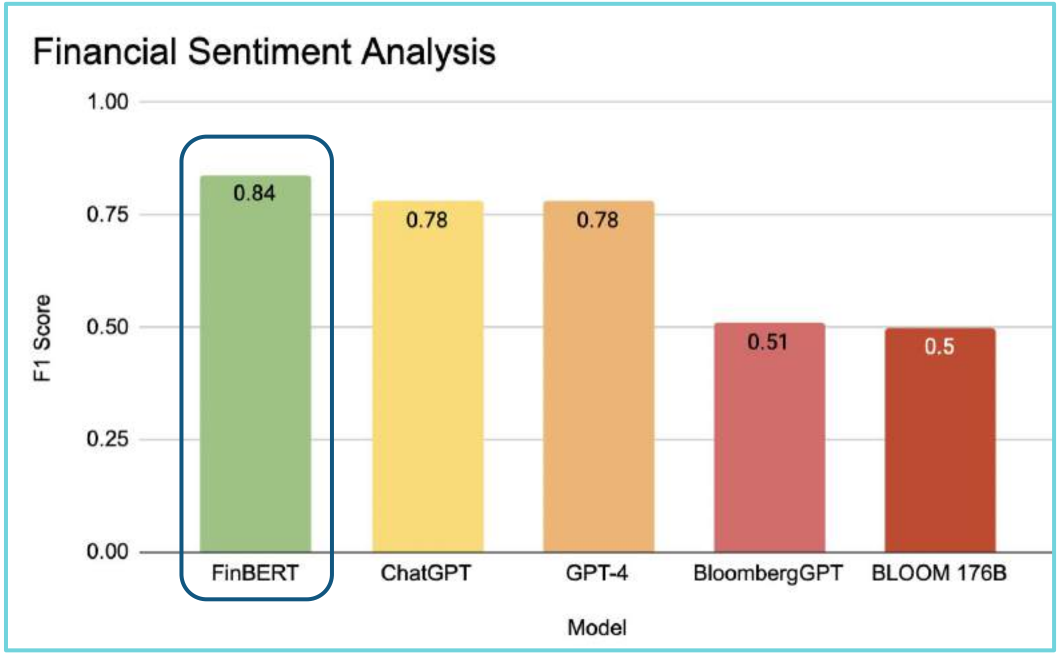


ChatGPT Wins 💪

https://arxiv.org/pdf/2304.04339.pdf
https://www.sciencedirect.com/science/article/pii/S156625352300177X

@rajistics

Compare the accuracy



FinBERT Wins 🤷🏻‍♀️

100M beats 1T

https://arxiv.org/pdf/2305.05862.pdf

FinBERT beats GPT-4

BioBERT beats GPT-4

SqlCoder beats GPT-4

Stockfish beats GPT-4

DeepL beats GPT-4

Specialists win!

https://github.com/zeno-ml/zeno-build/tree/main/examples/analysis_gpt_mt/report
https://arxiv.org/pdf/2303.17728.pdf
https://github.com/defog-ai/sqlcoder
https://chessily.com/blog/stockfish-vs-gpt-4-bing-ai/

Pretraining versus Fine Tuning a LLM

Pre-trained	Fine-Tuned
F1: 0.51	F1: 0.85
$2.5 million	$65

Pretrain/Foundation models is the last resort!

F1 for Financial Phrase Bank dataset
https://github.com/AI4Finance-Foundation/FinGPT

Why Specialist/Smaller Model?

Accuracy: a smaller model fine-tuned for a specific purpose will almost always outperform a larger general-purpose model

Speed: the smaller a model is, the faster it predicts

Cost: they're less expensive to train and host

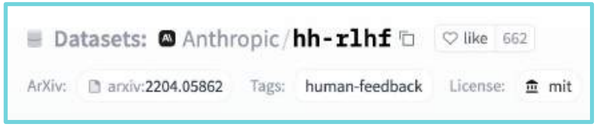
Explainability/MRM: they're easier to understand and test for risk

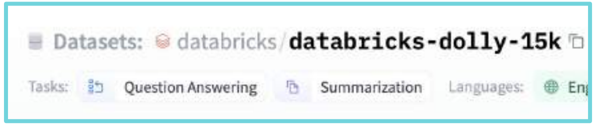
Agility: they're faster to train and retrain, letting you iterate quicker

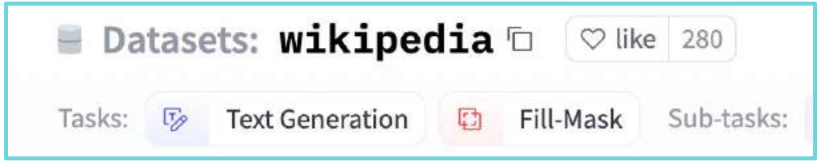
MLOps: established practices for managing smaller models

@rajistics

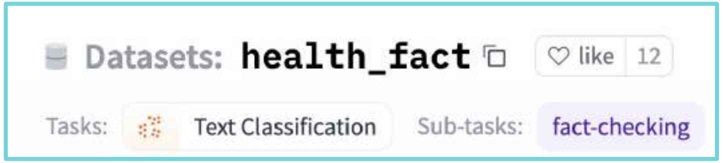
Specialist Datasets







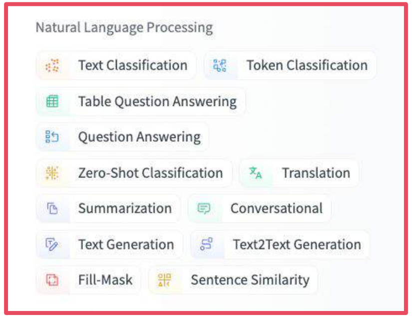
Hugging Face hub has over 60k datasets

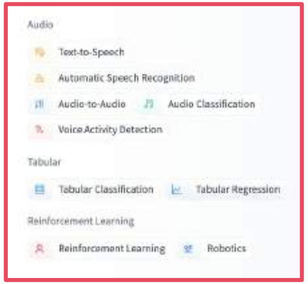


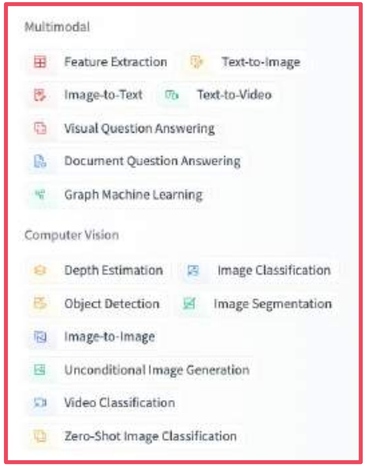
+ Domain (Finance, Healthcare, Environmental, Astronomy, . . .)

@rajistics

Specialist Models







+ Domain (Finance, Healthcare, Environmental, Astronomy, . . .)

Hugging Face hub has over 300k models

Open Source LLMs

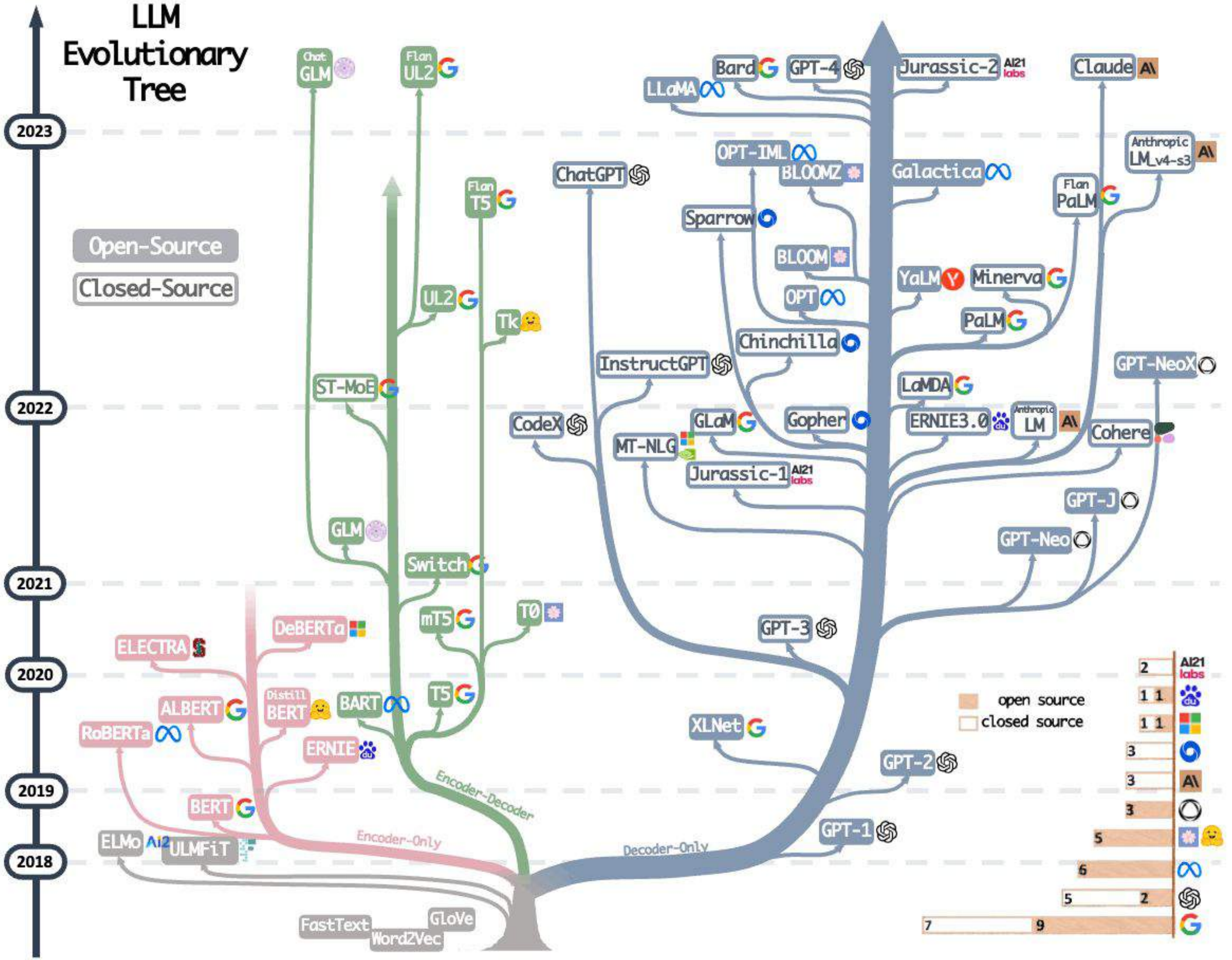
Trends: Text->Image Generation

Dalle-Mini
May 2022
30 seconds

Stable Diffusion
August 2022
8 seconds

Stable Diffusion XL
July 2023
4 seconds

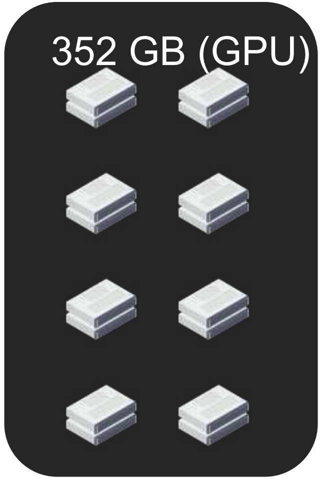
So many LLMs!



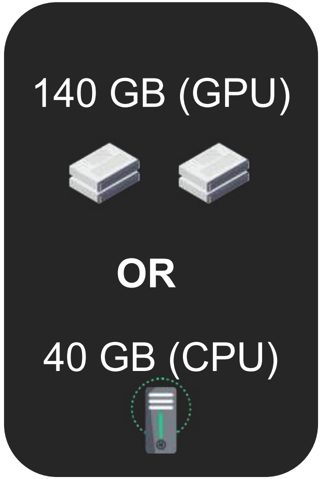
https://github.com/Mooler0410/LLMsPracticalGuide

@rajistics

Trends: Open Source LLMs



BLOOM (176B)
July 2022
MMLU: 39.13

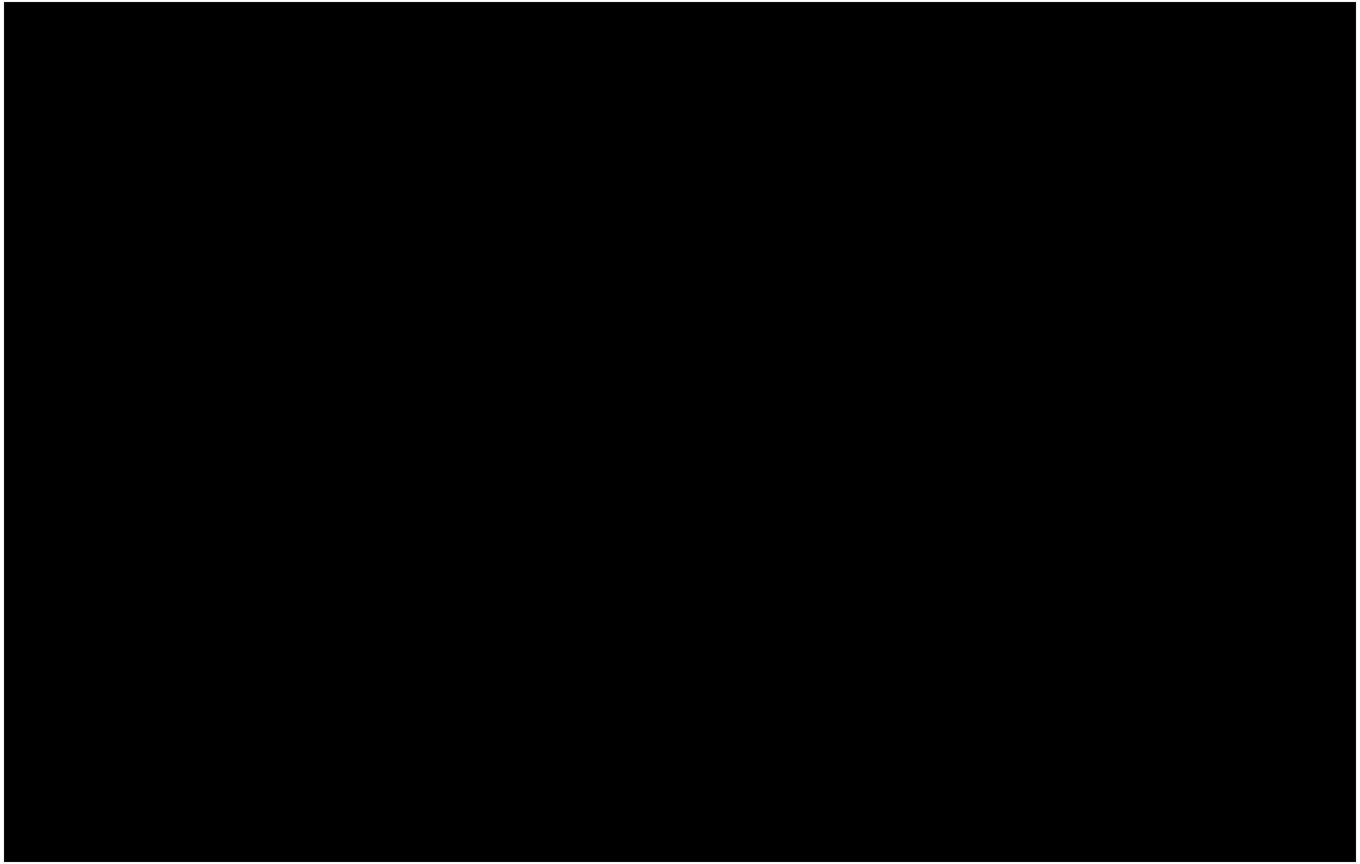


Llama-2 (70B)
August 2023
MMLU: 68.9

Smaug (72B)
February 2024
MMLU: 77

Open Source LLM Leaderboard

more than
1600 LLMs
evaluated!



Pretrained Models

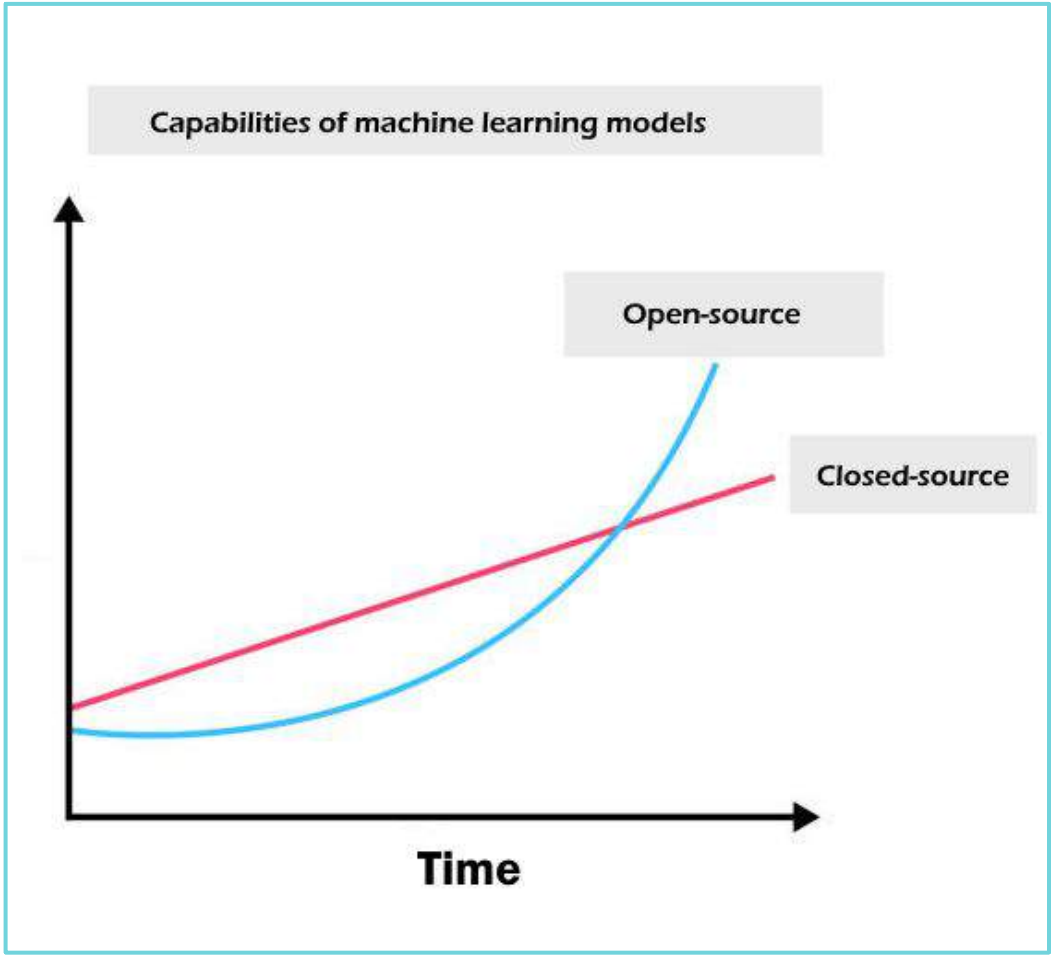
Open AI:
GPT-4 (8K)
GPT-4 (32K)
GPT-3.5 (4k)
GPT-3.5 (16k)
babbage-002
davinci-002

Open Source:
Falcon (180B)
LLama-2 (70B)
Tigerbot
LLama (65B)
Falcon (40B)
LLama-2 (13B)
MPT (30B)
Atom_GPT

Open Source:
Gowizard
Phi-1
Galactica
TinyStories
Palmyra-Large
RedPajama
GPT-NeoX

80 more

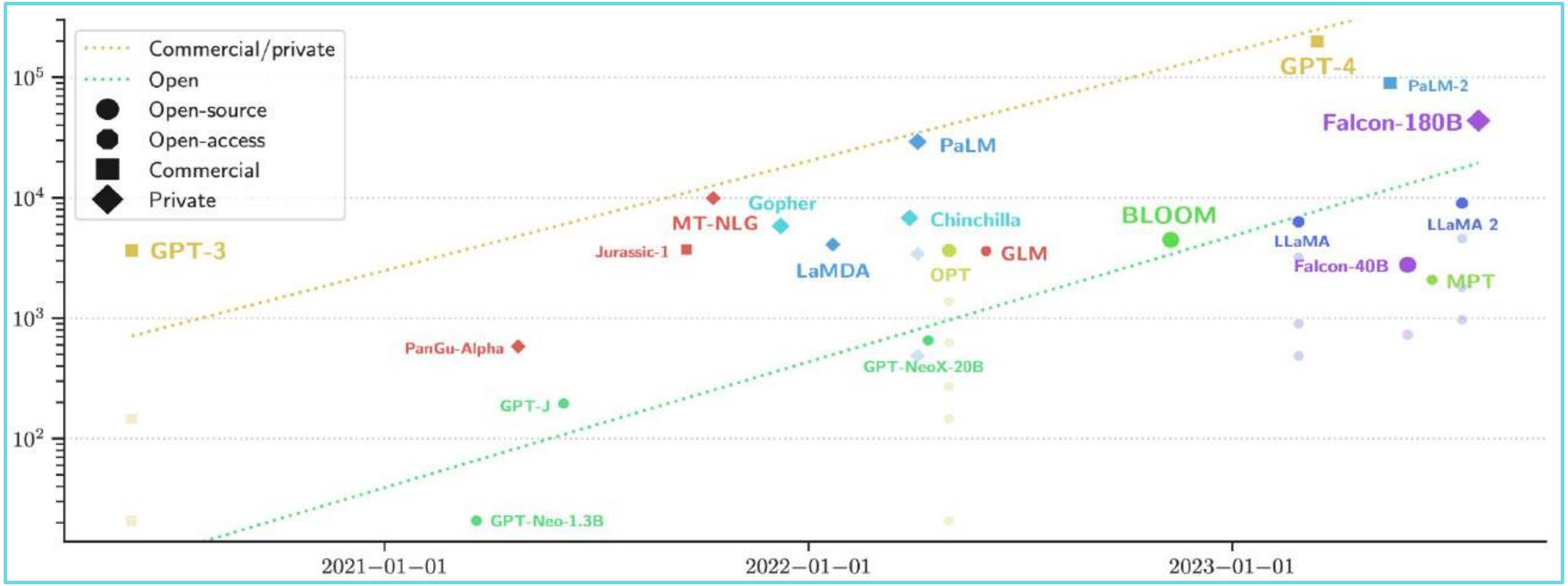
Alternative to closed source software



Abubakar Abid

@rajistics

Trends: OSS versus Commercial Compute

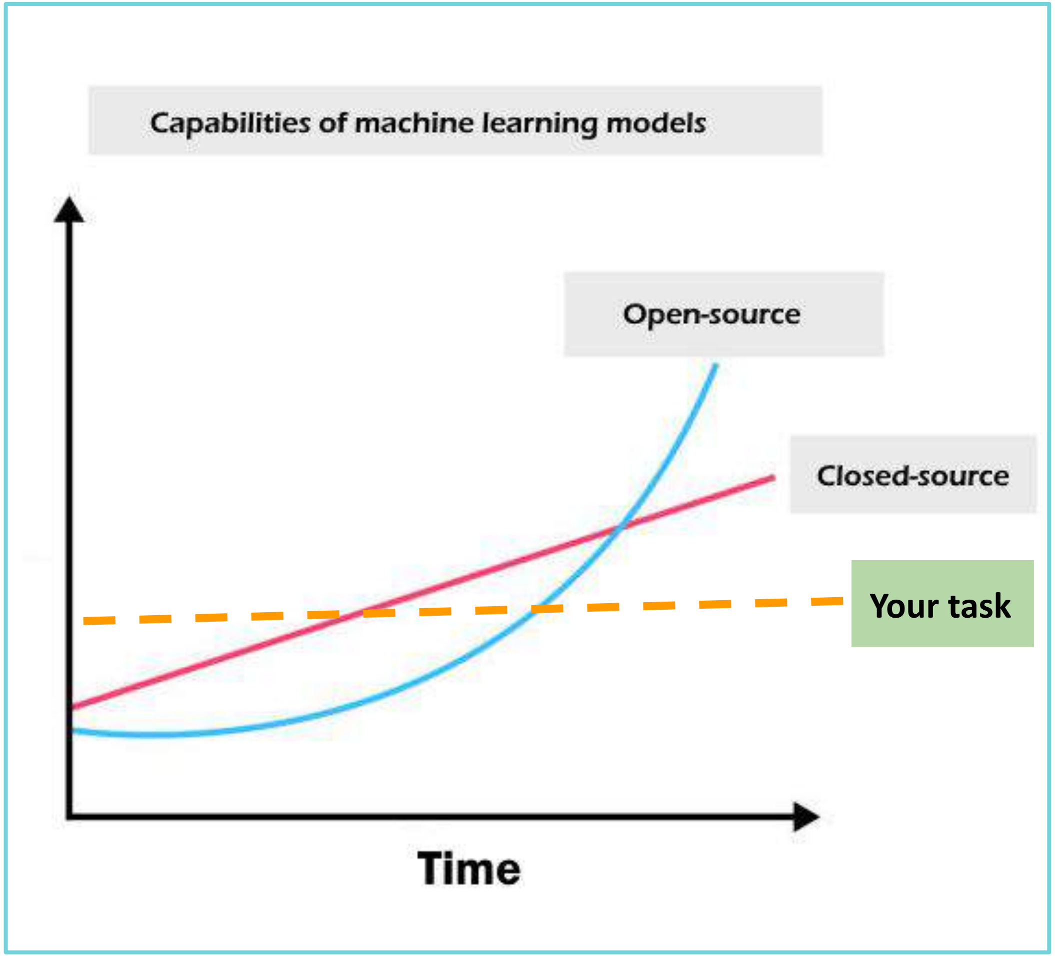


Open-source is lagging, but keeping pace

@rajistics

Julien Launay - Compute [PF-days]

Evaluate based on your task



Abubakar Abid

Trends: NPS versus popularity

Open-source is lagging in popularity and satisfaction

https://retool.com/reports/state-of-ai-2023

Why Open Source?

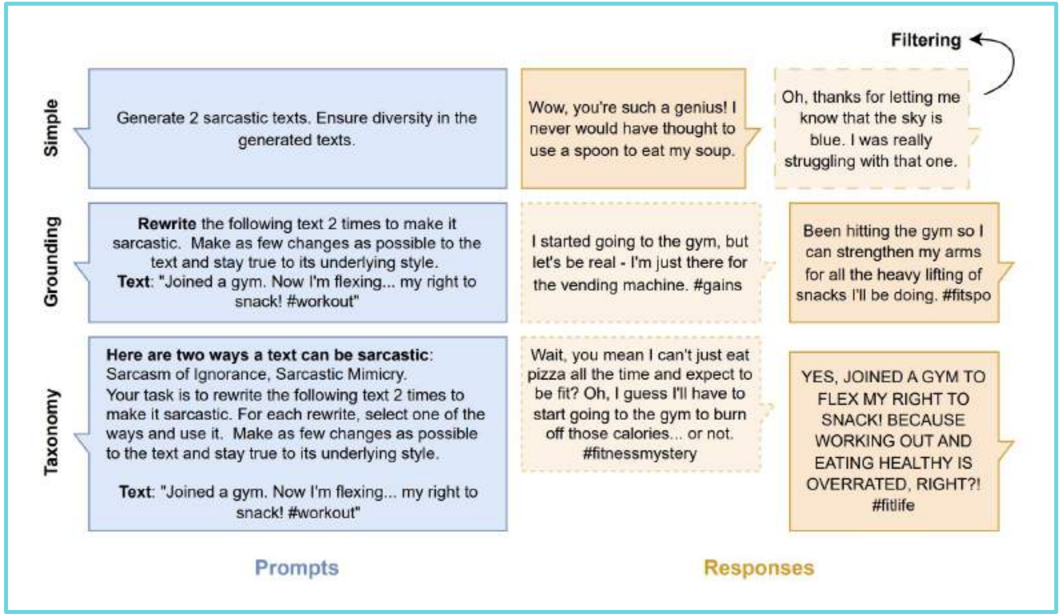
IP protection: customers train their models on their data, and own them.
Freedom of choice: customers are not locked in. They can switch models anytime
Privacy: customers don't have to send their data to black box APIs
Transparency: customers have full visibility on the model and the training data. They can better identify potential biases or errors
IT flexibility: customers can train and deploy models anywhere they like


quality
results
analysis
business
consistency
BEST
solutions
customers
success
standard
PRACTICE
technique
organiza
skills
management
quality
improvem
development
satisfactio
optimal
good

Generating Data with LLMs

LLM can create synthetic evaluation datasets for

- Pretraining
- instruction-tuning
- preference-tuning

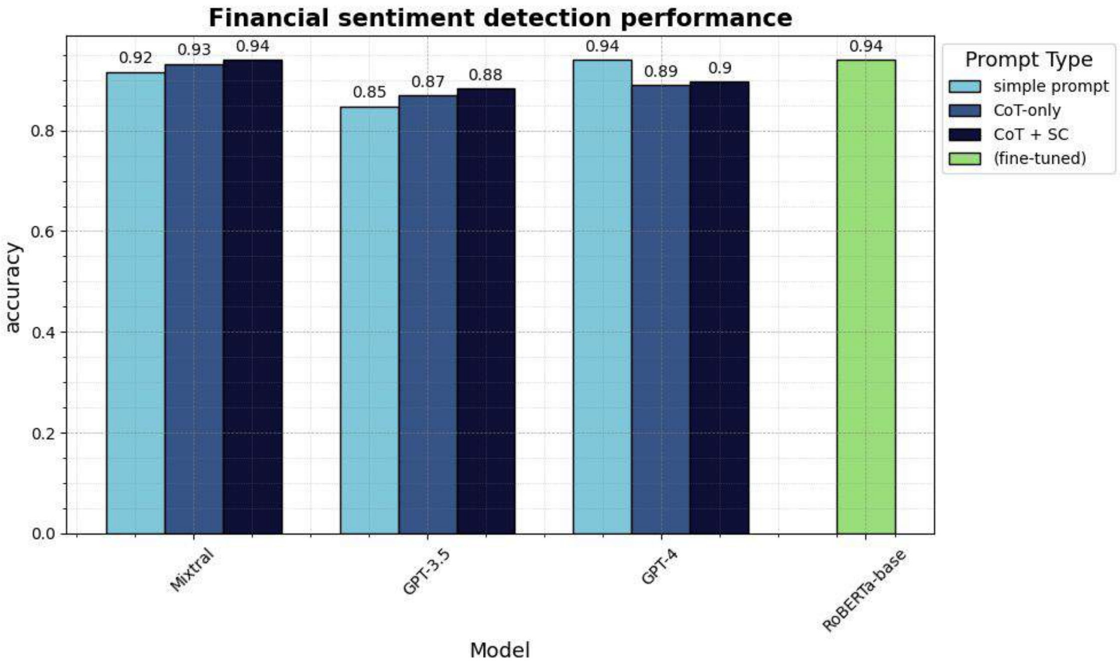


Generating Faithful Synthetic Data with Large Language Models: A Case Study in Computational Social Science
https://arxiv.org/pdf/2305.15041.pdf

Better Data -> Cheaper and Faster

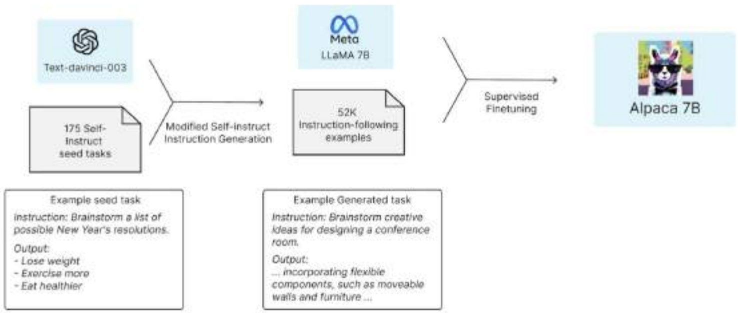
Cost to process 1 M sentences

- RobertA - $2.7
- GPT3.5 - $153

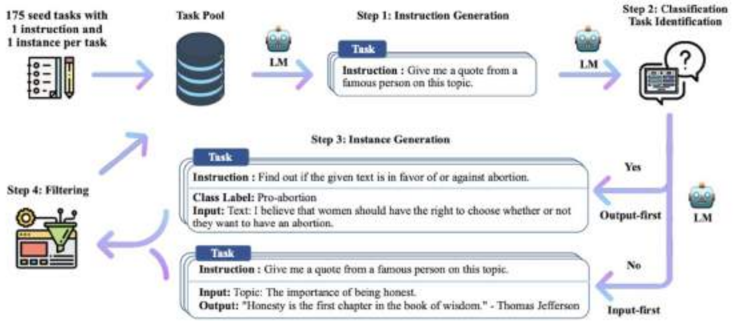


Synthetic data: save money, time and carbon with open source
https://huggingface.co/blog/synthetic-data-save-costs

Use Distillation or Self Improvement



Distillation



Self-Improvement

Synthetic Data for Finetuning: Distillation and Self-Improvement
https://eugeneyan.com/writing/synthetic/

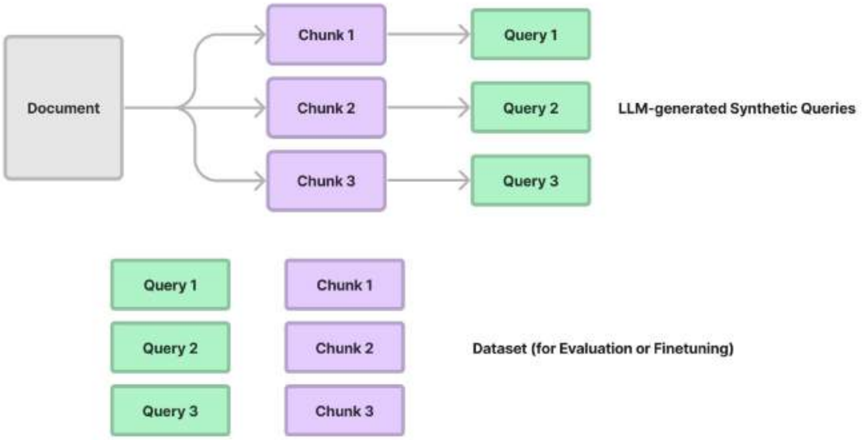
@rajistics

Pro Tip: Generate an synthetic evaluation dataset

You can use a LLM to help create synthetic evaluation datasets

Anthropic: https://github.com/anthropics/anthropic-cookbook/blob/main/long_context/mc_qa.ipynb

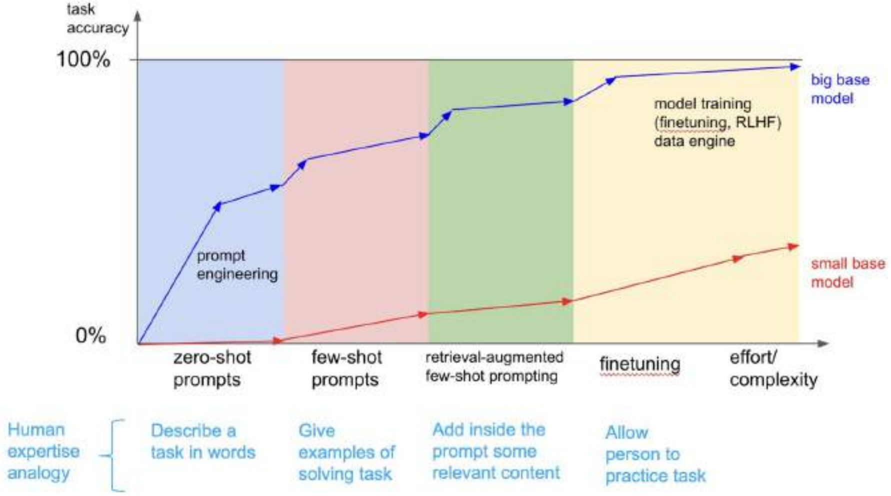
Llama-Index: https://gpt-index.readthedocs.io/en/v0.8.30/examples/low_level/evaluation.html



https://www.databricks.com/blog/LLM-auto-eval-best-practices-RAG
Jerry Liu: Evaluating and Optimizing your RAG App

Fine Tuning LLMs - Why

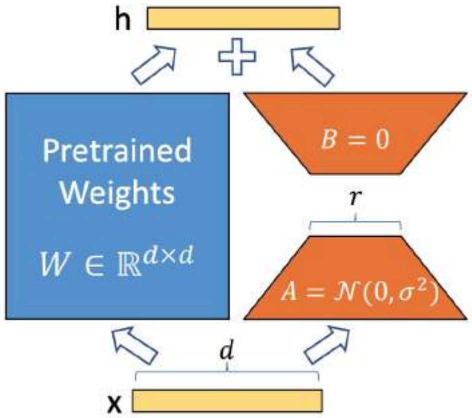
- Improve model performance
- Improve model efficiency (smaller)



Comparing LLM fine-tuning methods:
https://www.signalfire.com/blog/comparing-llm-fine-tuning-methods

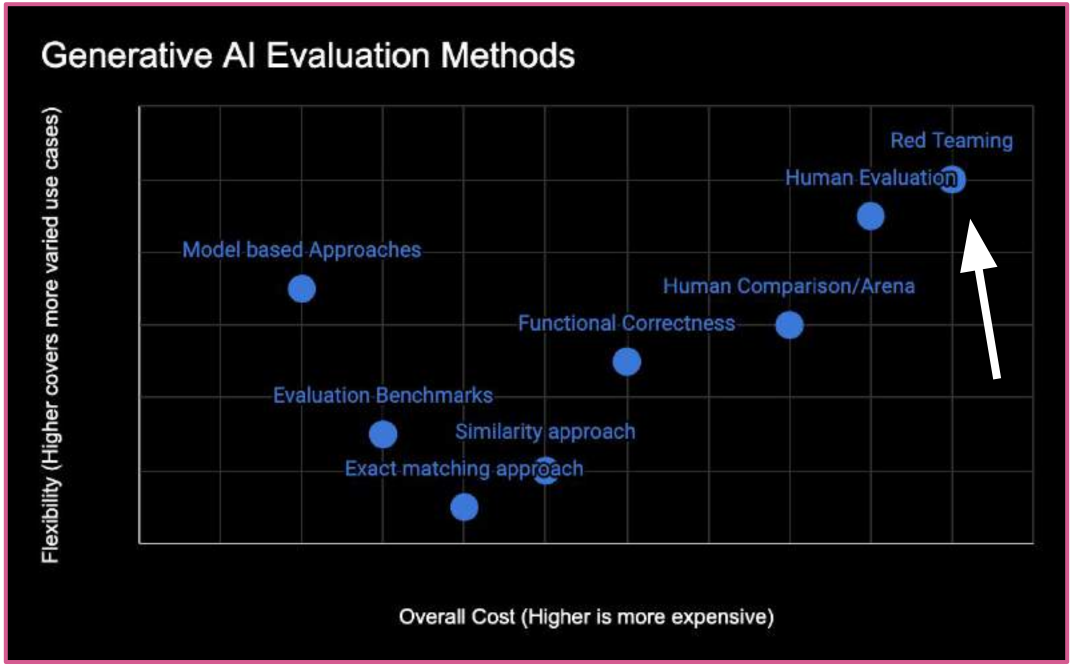
Fine Tuning LLMs - How

- Supervised Fine-Tuning
- Parameter-Efficient Fine-Tuning (PeFT)
  - LoRA



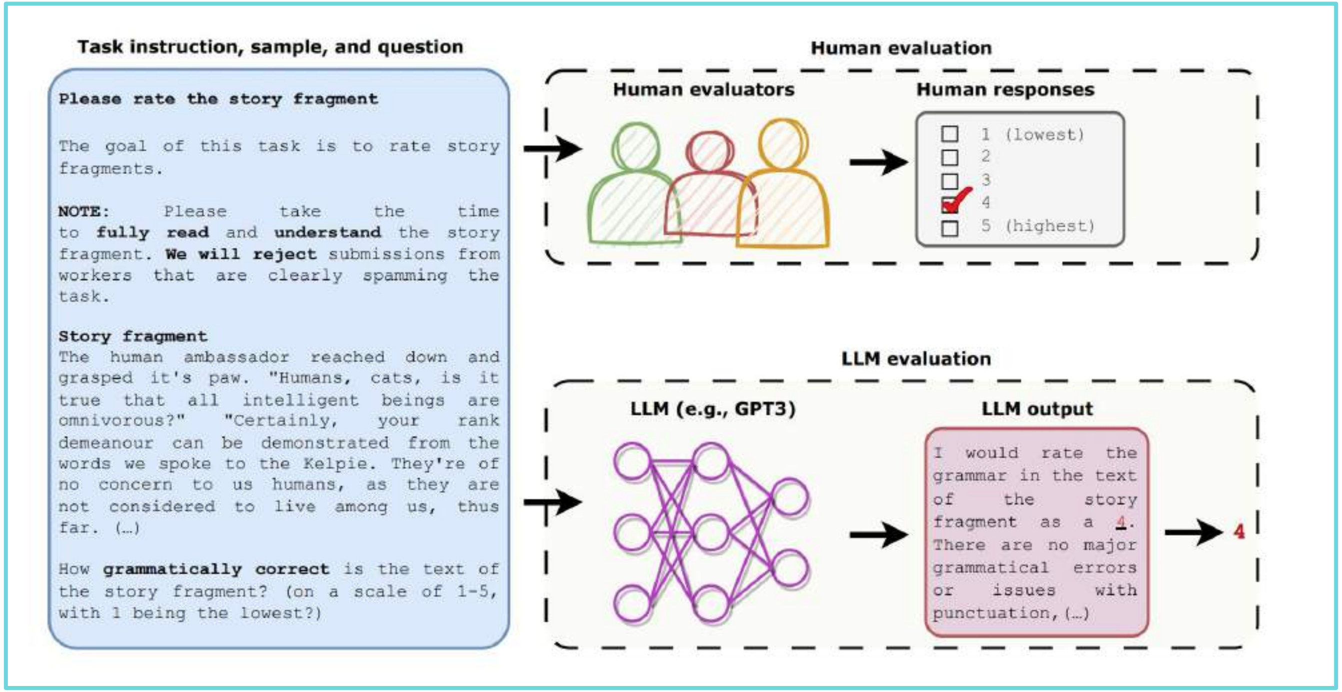
Methods for evaluating Generative AI

- Exact matching approach
- Similarity approach
- Functional Correctness
- Evaluation Benchmarks
- Human Evaluation
- Human Comparison/Arena
- Model based Approaches
- Red Teaming



@rajistics

Model based evaluation



C’mon Man - This isn’t going to work

Bharat Saxena • 1st 2d •••

Bringing intelligence to Mainframes @ BMC Software | Explainable AI (XAI) | NLP ...

Rajiv Shah From personal experience, I am a big skeptic when it comes to using another model as an evaluator ... Hopefully you will be able to share some details from your presentation as some time in future.

Bright lines for model based evaluation

Assertion/Condition

- Length
- Language Match

Well known problems

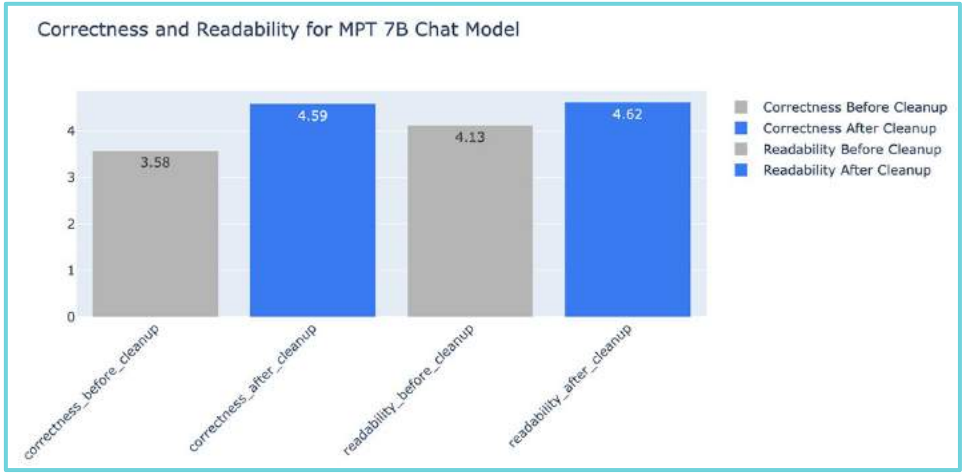
- Sentiment
- Toxicity

These evaluation prompts that take very little judgement on behalf of the model as an evaluator

Results: Improving Data Quality

Data cleaning improved the **correctness** of the LLM generated answers by up to **+20%**

Cleaning also **reduced** the number of tokens for the context by up to **-64%**



https://www.databricks.com/blog/announcing-mlflow-28-llm-judge-metrics-and-best-practices-llm-evaluation-rag-applications-part

Model based evaluation - Professionalism

📄 Define Professionalism

🎹 Grading Scale

🤖 Select a model

```
professionalism = mlflow.metrics.make_genai_metric(
    name="professionalism",
    definition=(
        "Professionalism refers to the use of a formal, respectful, and appropriate sty
        "tailored to the context and audience. It often involves avoiding overly casual
        "colloquialisms, and instead using clear, concise, and respectful language."
    ),
    grading_prompt=(
        "Professionalism: If the answer is written using a professional tone, below are
        "- Score 1: Language is extremely casual, informal, and may include slang or co
        "professional contexts."
        "- Score 2: Language is casual but generally respectful and avoids strong infor
        "some informal professional settings."
        "- Score 3: Language is overall formal but still have casual words/phrases. Bor
        "- Score 4: Language is balanced and avoids extreme informality or formality. Su
        "- Score 5: Language is noticeably formal, respectful, and avoids casual element
        "business or academic settings. "
    ),
    examples=[professionalism_example_score_1, professionalism_example_score_2, profess
    model="openai:/gpt-4",
    parameters={"temperature": 0.0},
    aggregations=["mean", "variance"],
    greater_is_better=True,
)
```

https://www.databricks.com/blog/announcing-mlflow-28-llm-judge-metrics-and-best-practices-llm-evaluation-rag-applications-part

Model based evaluation - Professionalism

📄 Define Professionalism

🎹 Grading Scale

🤖 Select a model

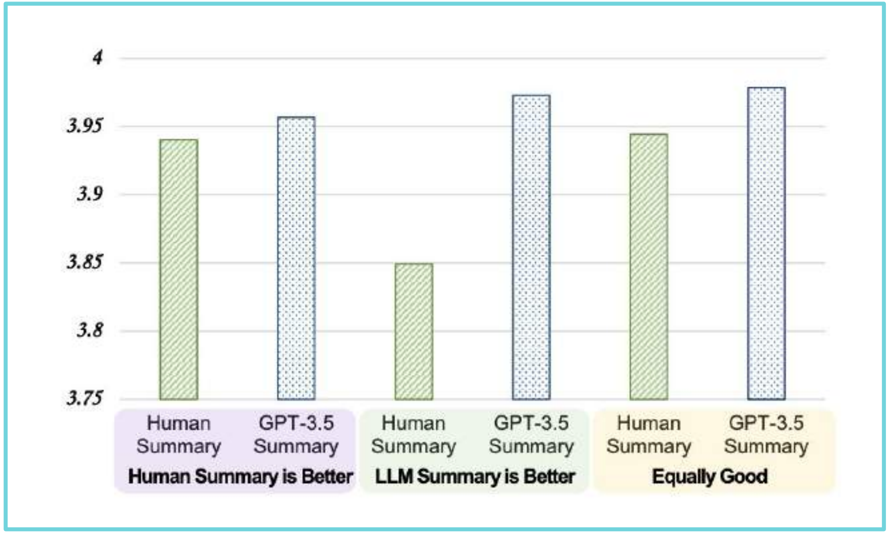
```
professionalism_example_score_2 = mlflow.metrics.EvaluationExample(
    input="What is MLflow?",
    output=(
        "MLflow is like your friendly neighborhood toolkit for managing your machine le
        "you track experiments, package your code and models, and collaborate with your
        "workflow smoother. It's like your Swiss Army knife for machine learning!"
    ),
    score=2,
    justification=(
        "The response is written in a casual tone. It uses contractions, filler words s
        "exclamation points, which make it sound less professional. "
    ),
)
```

https://www.databricks.com/blog/announcing-mlflow-28-llm-judge-metrics-and-best-practices-llm-evaluation-rag-applications-part

Model evaluation – human alignment

It appears to align with humans

Human and GPT-4 judges can reach above 80% agreement on the correctness and readability score. And if we lower the requirement to be smaller or equal than 1 score difference, the agreement level can reach above 95%.



https://arxiv.org/abs/2305.01937
https://www.databricks.com/blog/LLM-auto-eval-best-practices-RAG
https://arxiv.org/abs/2303.16634
https://arxiv.org/pdf/2306.05685.pdf

Summary: Model based evaluation

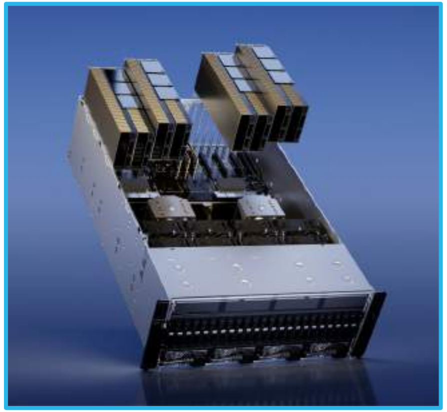
✅ Cheaper and faster than human evaluation

✅ Align better with humans than reference-based and reference free baselines

✅ Can provide a more fine grained continuous score by re-weighting the discrete scores by their respective token probabilities.

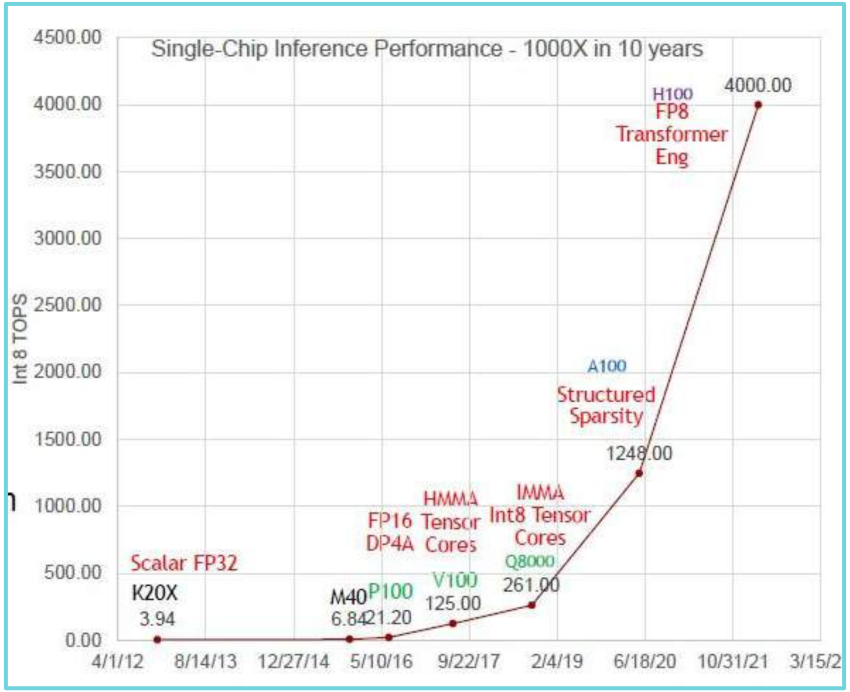
❌ Sensitive to the instructions and prompts.

❌ Several known biases



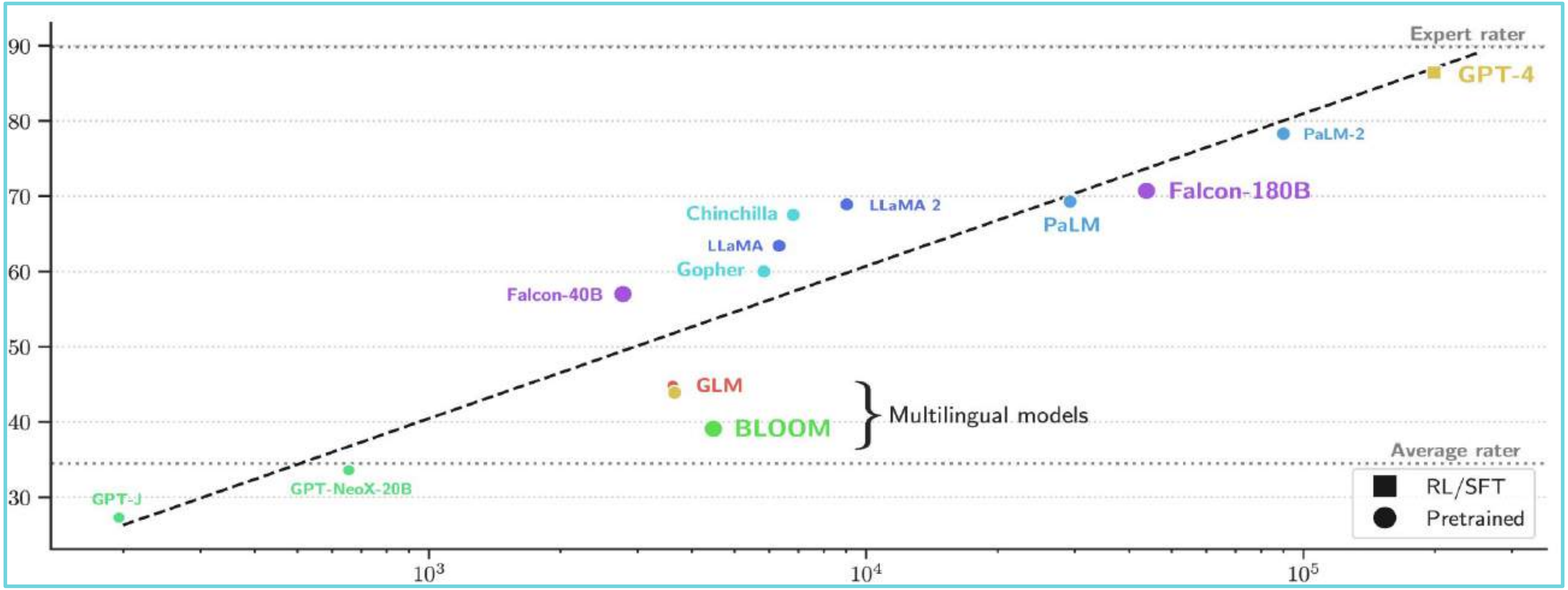
Resources for Generative AI

Trends: 1000X on Compute



https://epochai.org/blog/who-is-leading-in-ai-an-analysis-of-industry-ai-research

Trends: Knowledge versus compute

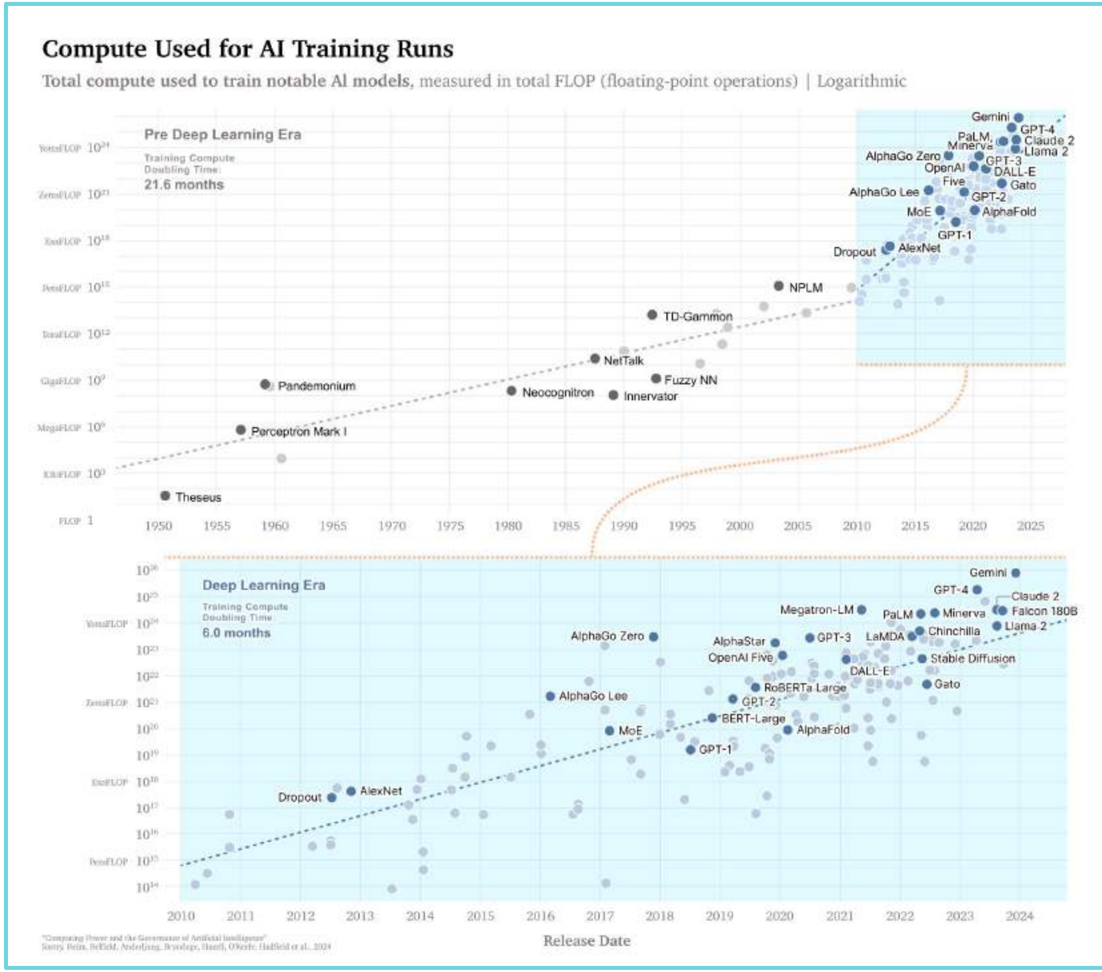


Model performance is predictable!

Julien Launay - MMLU 5-shot [acc. %]

@rajistics

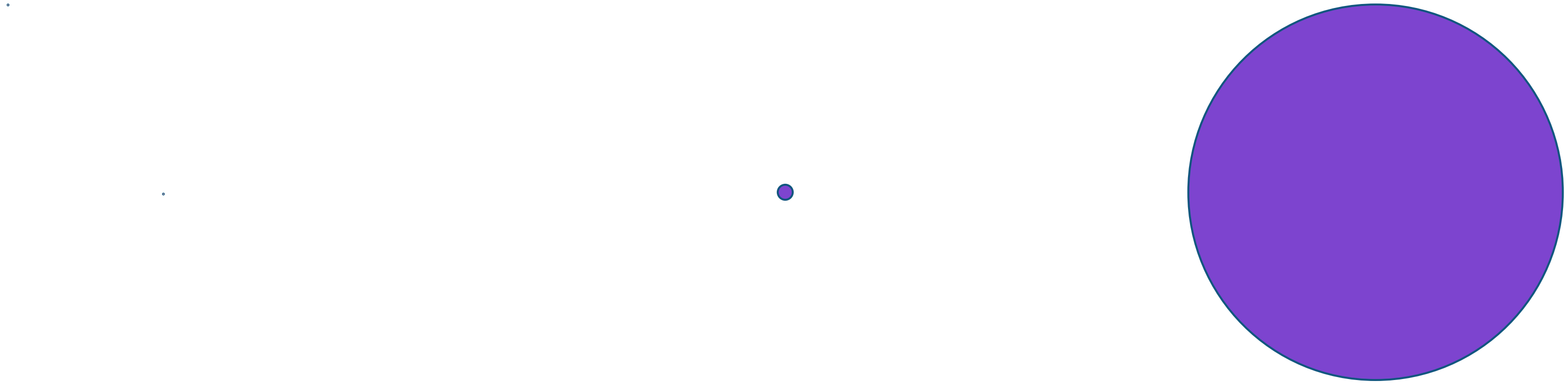
Trends: Compute



Models get better with more compute

Computing Power and the Governance of Artificial Intelligence: https://arxiv.org/pdf/2402.08797.pdf

Trends: Compute required for LLMs



DistilBERT
60M Parameters
48.5 GFLOPs/query

Llama-2
7B parameters
15 TFLOPs/query

GPT-4
111B * 16 parameters
~600 TFLOPs/query

** - Guesses by Raj, don't plan on it

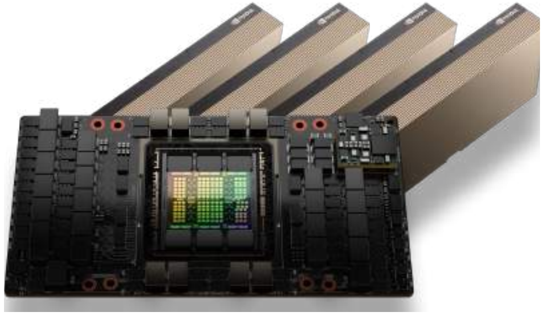
Trends: Compute Options for LLMs

iPhone 8
400 GFLOPs

DistilBERT
60M Parameters
48.5 GFLOPs/query

iPhone 14
2 TFLOPs

Llama-2
7B parameters
15 TFLOPs/query



NVIDIA H100
67 TFLOPs

GPT-4
111B * 16 parameters
~600 TFLOPs/query

** - Guesses by Raj, don't plan on it

@rajistics

Trends: Access to Data is Harder

Twitter/X

Reddit

AP News

Stack Overflow

Widespread
scraping for data

Paid licensing for data
Opt out of training in robots.txt

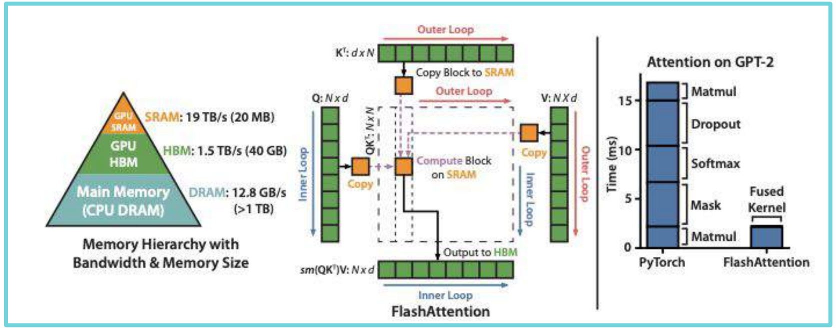
Trends: Access to Data is Easier

Hugging Face Hub

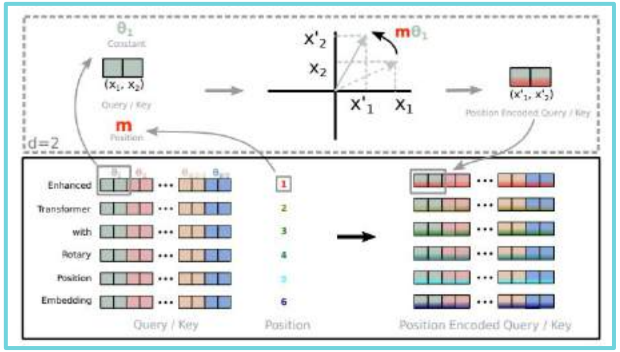
Organizations sharing datasets

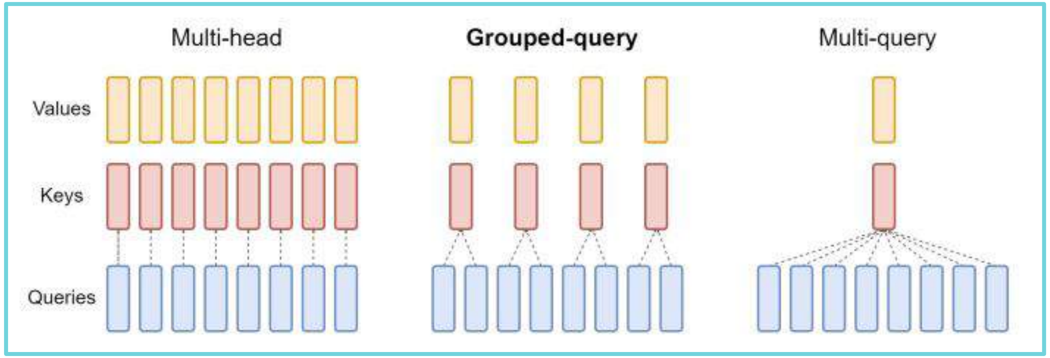
Recent Architectural Improvements

Rotary Position Embedding (RoPE)



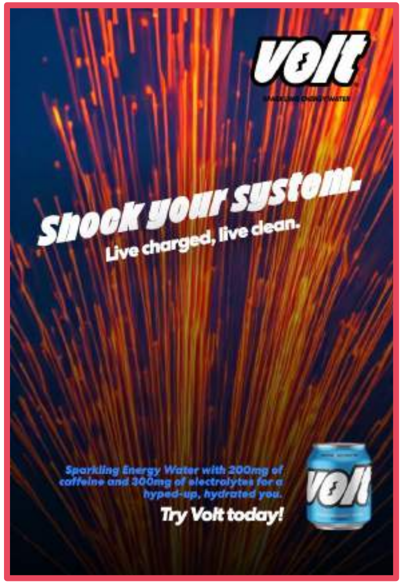
Flash Attention

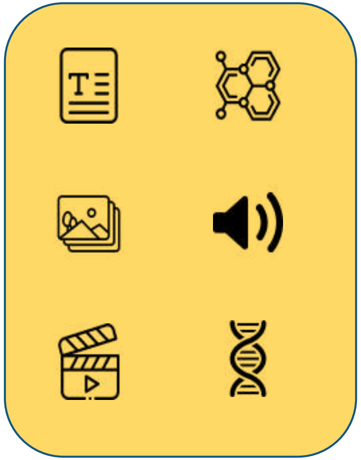




Flash Attention:https://arxiv.org/abs/2205.14135
ROPE: https://arxiv.org/abs/2104.09864
MQA: https://arxiv.org/pdf/2305.13245.pdf

Trends: MultiModal





Moving to production: LayoutLM, GPT4, IDEFICS

https://www.newscientist.com/article/2374607-ai-passed-an-advertising-turing-test-for-the-first-time

@rajistics

Risks with LLMs: Hallucinations

Generative models are always dreaming

The New York Times

Here's What Happens When Your Lawyer Uses ChatGPT

A lawyer representing a man who sued an airline relied on artificial intelligence to help prepare a court filing. It did not go well.

https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html

Risks of Large Language Models

Bias: model predictions that favor particular groups
Untrue outputs / Hallucinations: models quite confidently output false information
Interpretability: don’t have good tools to understand these models
Legal concerns: did you license the training data for the model? are the outputs of the model infringing?
Security: new attacks like prompt injection

Trends: Impact of Jobs

AI Providing Ethical Advice

AI for story design and acting

Negotiating
a NDA

https://www.gamesradar.com/from-star-wars-to-starfield-voice-actors-hit-out-at-microsofts-ai-decision-if-you-want-to-start-a-voice-acting-career-dont-bother/
https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2023/10/Can-AI-Provide-Ethical-Advice_2.pdf
https://www.bbc.com/news/business-67238386

Generative AI:
A Survey of Current Practices, Challenges, and Best Practices

Rajiv Shah
@rajistics
r.shah@snowflake.com